



## 【特許請求の範囲】

【請求項1】 情報記録媒体又は情報通信網から供給された情報が記憶される記憶部と、該記憶部の情報を類似する情報でまとめて代表情報と従属情報を決める類似判定手段と、前記記憶部の情報を提示する場合に類似する情報については代表情報を提示する提示手段と、を具備する情報フィルタ装置。

【請求項2】 記憶部は、情報に割り振られた複数のキーワードと過去にそれらキーワードの割り振られた情報を必要／不要とした履歴とに基づいて計算された必要性の高さに応じて情報が並べられていることを特徴とする請求項1記載の情報フィルタ装置。

【請求項3】 類似判定手段は、記憶部に記憶された各情報に付けられているキーワードの一致度に基づいて代表情報に類似する情報を検索することを特徴とする請求項1又は請求項2記載の情報フィルタ装置。

【請求項4】 提示手段は、代表情報及び独立情報を必要性の高い順序で提示することを特徴とする請求項1乃至請求項3のいずれかに記載の情報フィルタ装置。

【請求項5】 提示手段は、代表情報を提示する際に、当該代表情報に類似した従属情報の数又は有無を一緒に表示することを特徴とする請求項1乃至請求項4のいずれかに記載の情報フィルタ装置。

【請求項6】 類似判定手段は、記憶部の情報を必要性の高い情報から順番にチェックして未だ代表情報又は従属情報になっていない独立情報があれば、その独立情報を代表情報とし類似する従属情報を当該代表情報から必要性の低い方向に検索することを特徴とする請求項1乃至請求項5のいずれかに記載の情報フィルタ装置。

【請求項7】 類似判定手段は、類似する情報及び独立情報を検索する検索範囲が予め指示又は設定され、当該検索範囲を必要性の低い方向へ順次シフトさせることを特徴とする請求項6記載の情報フィルタ装置。

【請求項8】 情報に割り振られた複数のキーワード信号をベクトル信号に変換する手段と、必要な情報から構成される肯定メトリック信号、不要な情報から構成される否定メトリック信号及び前記ベクトル信号を用いて肯定スコア及び否定スコアを計算するスコア計算部と、前記肯定スコア信号及び前記否定スコア信号からなる二次元平面の点の分布を必要な情報及び不要な情報に分離する判定パラメータ信号、前記肯定スコア信号及び前記否定スコア信号から必要性信号及び信頼性信号を計算する必要性計算部と、情報を前記必要性信号の大ききの順に並べて前記記憶部に書き込む手段とを具備する請求項1乃至請求項7のいずれかに記載の情報フィルタ装置。

【請求項9】 情報記録媒体又は情報通信網から供給された情報を類似する情報でまとめて代表情報と従属情報を決め、前記情報を提示する場合に類似する情報については代表情報を提示することを特徴とする情報フィルタリング方法。

【請求項10】 必要性の高さに応じた順番で情報が並べられた記憶部から情報を取り出し、情報に付けられているキーワードの一致度に基づいて代表情報に類似する従属情報を検索し、類似する情報については代表情報を提示対象として必要性の高い順序で情報を提示することを特徴とする請求項9記載の情報フィルタリング方法。

【請求項11】 コンピュータによって情報をフィルタリングする情報フィルタリングプログラムを記録した記録媒体であって、情報フィルタリングプログラムはコンピュータに、情報記録媒体又は情報通信網から供給された情報から代表情報に類似する従属情報を検出させ、類似する情報については代表情報を提示対象として必要性の高い順序で情報を提示させる情報フィルタリングプログラムを記録した記録媒体。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、電子または光等を媒体とする記憶装置や情報通信網から必要な情報を取り出し易くする情報フィルタ装置及び情報フィルタリング方法並びに情報フィルタリングプログラムを記録した記録媒体に関するものである。

## 【0002】

【従来の技術】近年の情報通信基盤の発展に伴う情報通信網の大規模化及び通信量の著しい増大に対処するための技術として、情報フィルタ装置の実現が強く望まれている。この背景には、今日、個人が処理可能な情報量に対して、個人がアクセスできる情報量が上回るようになってきていることがある。このために、大量の情報の中にユーザにとって必要な情報が埋没することが、しばしば起こる。

【0003】情報フィルタ装置に関連する従来技術としては、特許検索などに用いられるキーワード論理式をあげることができる。すなわち、数十万から数百万件に及ぶ特許情報をキーワード論理式によりフィルタリングするものである。

【0004】しかしながら、キーワード論理式を用いる従来の検索においては、ユーザがキーワードについての論理式を精度良く設定する必要があるので、ユーザがファインリングされているデータ群の癖（例えば、どのような条件の基に、当該データのキーワードが決定されているのか等）やシステムの構造（例えば、キーワードがシソーラス体系のあるシステムであるか否か等）を十分に知り得ていなければ良い検索ができない。このため、初心者には精度の高い情報フィルタリングを行うことができないという課題があった。

【0005】また、情報フィルタリングした結果はキーワードについての論理式に適合するという評価があるだけであり、たまたまキーワードでは合致しているが、内容は求めているものとは異なるケースがあったり、あるいは多くの検索結果からユーザにとって必要度の高い情

報をその結果から順に取り出すことは容易ではない。

【0006】そこで、本発明者らは、初心者にも精度の高い情報フィルタリングが可能で、ユーザにとって必要性の高い情報を容易に取り出すことのできる情報フィルタ装置を開発し、特願平8-230012号及び特願平9-046384号として特許出願済みである。

【0007】この情報フィルタ装置は、情報に割り振られた複数のキーワードをベクトルに変換し、ベクトルとユーザがどんな情報を必要とし不要としたかを表現した行列を用いてスコアを計算して、情報の必要性和信頼性を計算する。必要性の大きい順に情報を並べ換え、ユーザから要求があれば必要性の大きい順に情報を提示する。さらに、提示した情報が必要か不要かというユーザの評価を入力して、ユーザの評価と前記複数のキーワードとからスコア計算にもちいる行列を修正するものである。

【0008】この構成によって、複数のキーワードは、距離の定義ができない記号から、ユーザの必要度を反映したメトリックを用いて距離を定義できるベクトル表現へと変換され、ユーザの必要度を定量化することができ、ユーザは必要性の高い情報から順に情報を得ることができるようになる。

【0009】

【発明が解決しようとする課題】しかしながら、上記情報フィルタ装置においては、複数の情報源（例えば、複数の新聞社）から情報を集める場合、同じ内容の情報が繰り返し現れ、十分な情報の一覧性が得られないという課題があった。

【0010】例えば、多くの新聞紙面で報じられるような重大事故が発生した場合、ユーザにとっても興味深いその重大事故（情報）に高いスコアが付けられることになる。その結果、必要性の大きい順に情報を提示すると、情報提示内容がその重大事故（情報）の記事で占有され、その他の情報を得るまでに同じ内容の記事が何度も繰り返し提示されることになる。

【0011】本発明は、以上のような実情に鑑みてなされたものであり、同じような情報が複数含まれる場合には、代表的な情報と関連情報の件数表示することにより、できる限り多くの種類の情報を一度に表示でき一覧性の高い情報フィルタ装置及び情報フィルタリング方法を提供することを目的とする。

【0012】

【課題を解決するための手段】上記課題を達成するために、本発明は以下のような手段を講じた。

【0013】請求項1記載の発明は、情報記録媒体又は情報通信網から供給された情報が記憶される記憶部と、該記憶部の情報を類似する情報でまとめて代表情報と従属情報を決める類似判定手段と、前記記憶部の情報を提示する場合に類似する情報については代表情報を提示する提示手段とを具備する構成を採る。

【0014】請求項9記載の発明は、情報記録媒体又は情報通信網から供給された情報を類似する情報でまとめて代表情報と従属情報を決め、前記情報を提示する場合に類似する情報については代表情報を提示する構成を採る。

【0015】本発明によれば、記憶部の情報を提示する場合に類似する情報については代表情報を提示するので、多くの種類の情報を一度に表示でき一覧性の改善を図ることができる。

【0016】請求項2記載の発明は、記憶部の情報を、情報に割り振られた複数のキーワードと過去にそれらキーワードの割り振られた情報を必要／不要とした履歴とに基づいて計算された必要性の高さに応じた順序で並べる構成を採る。

【0017】本発明によれば、記憶部の情報が必要性の高さに応じた順序で並べられているので、ユーザの好みに応じた順番で情報を提示できるとともに、類似する情報については代表する代表情報だけを表示して一覧性の改善を図ることができる。

【0018】請求項3記載の発明は、類似判定手段が、記憶部に記憶された各情報に付けられているキーワードの一致度に基づいて代表情報に類似する情報を検索する構成を採る。

【0019】本発明によれば、キーワードの一致度に基づいて代表情報に類似する情報を判断することができ、記憶部の情報を提示する場合に類似する情報については代表情報を提示するので、多くの種類の情報を一度に表示でき一覧性の改善を図ることができる。

【0020】請求項4記載の発明は、代表情報を必要性の高い順序で提示する構成を採る。

【0021】請求項10記載の発明は、必要性の高さに応じた順番で情報が並べられた記憶部から情報を取り出し、情報に付けられているキーワードの一致度に基づいて代表情報に類似する従属情報を検索し、類似する情報については代表情報を提示対象として必要性の高い順序で情報を提示する構成を採る。

【0022】本発明によれば、代表情報を必要性の高い順序で提示するので、ユーザの好みに応じた順番で情報を提示でき、類似する情報については代表する代表情報だけを表示して一覧性の改善を図ることができる。

【0023】請求項5記載の発明は、代表情報を提示する際に、当該代表情報に類似した従属情報の数又は有無を一緒に表示する構成を採る。

【0024】本発明によれば、代表情報と共に従属情報の数又は有無を表示するので、一覧性を確保しつつより多くの関連情報を知ることができる。

【0025】請求項6記載の発明は、記憶部の情報を必要性の高い情報から順番にチェックして未だ代表情報又は従属情報になっていない独立情報があれば、その独立情報を代表情報とし類似する従属情報を当該代表情報か

ら必要性の低い方向に検索する構成を採る。

【0026】本発明によれば、情報が重要性の高い順序で並んでいるので、代表情報よりも上に類似情報が存在しないことを利用して、類似情報を検索する際の検索範囲の効率化を図ることができる。

【0027】請求項7記載の発明は、類似する情報及び独立情報を検索する検索範囲が予め指示又は設定され、当該検索範囲を必要性の低い方向へ順次シフトさせる構成を採る。

【0028】本発明によれば、情報が重要性の高い順序で並んでいる場合に検索範囲を定めると、類似する情報は重要性が近似する特性が有り比較的近い位置に存在することを利用して、検索の効率化を図ることができる。

【0029】請求項8記載の発明は、情報に割り振られた複数のキーワード信号をベクトル信号に変換する手段と、必要な情報から構成される肯定メトリック信号、不要な情報から構成される否定メトリック信号及び前記ベクトル信号を用いて肯定スコア及び否定スコアを計算するスコア計算部と、前記肯定スコア信号及び前記否定スコア信号からなる二次元平面の点の分布を必要な情報及び不要な情報に分離する判定パラメータ信号、前記肯定スコア信号及び前記否定スコア信号から必要性信号及び信頼性信号を計算する必要性計算部と、情報を前記必要性信号の大きさの順に並べて前記記憶部に書き込む手段とを具備する構成を採る。

【0030】本発明によれば、複数のキーワードは、距離の定義ができない記号から、ユーザの必要度を反映したメトリックを用いて距離を定義できるベクトル表現へと変換され、ユーザの必要度を定量化することができ、ユーザは必要性の高い情報から順に情報を得ることができるようになる。

【0031】請求項11記載の発明は、コンピュータによって情報をフィルタリングする情報フィルタリングプログラムを記録した記録媒体であって、情報フィルタリングプログラムはコンピュータに、情報記録媒体又は情報通信網から供給された情報から代表情報に類似する従属情報を検出させ、類似する情報については代表情報を提示対象として必要性の高い順序で情報を提示させる情報フィルタリングプログラムを記録する構成を採る。

【0032】本発明によれば、類似の情報は代表的な情報にまとめられ一覧性の高い情報フィルタリングができるようになる。

【0033】

【発明の実施の形態】以下、本発明の実施の形態について図面を参照して詳細に説明する。

【0034】（実施の形態1）図1は実施の形態1に係る情報フィルタ装置の構成を示すブロック図であり、図2は情報フィルタ装置を機能単位にまとめたブロック図である。

【0035】この情報フィルタ装置は、ユーザーが過去

にどんな「情報」を必要としたかという履歴に基づいて「情報」のフィルタリングを行う情報フィルタリングユニット50、フィルタリングした情報をユーザーに提示するディスプレイ等のインタフェースユニット51、ユーザーがどんな「情報」を必要としたかという履歴に関する学習を行う学習ユニット52、フィルタリングした未読データ（ユーザーが未だ読んでいない情報）から内容がよく似た未読データにまとめて各グループを代表する「情報」の一覧にする代表情報一覧ユニット53に分けられる。

【0036】情報フィルタリングユニット50は、個々の「情報」につけられた複数のキーワード（正確には、分類コードを含む文字列）をベクトルに変換する部分と、ユーザーがどんな「情報」を必要／不要としたという履歴を表現した肯定メトリック信号及び否定メトリック信号を用いてある種のスコアを表す肯定信号と否定信号を計算する部分と、この肯定信号と否定信号とから「情報」の必要性をよく反映する必要性信号を計算する部分と、この必要性信号の大きい順に情報を並べ替える部分からなる。

【0037】図1では、ベクトル生成部1が「情報」に付けられたキーワードなどの複数の文字列をベクトルに変換し、符号辞書記憶部2がキーワードなどの複数の文字列をベクトルに変換するための符号辞書信号を記憶している。この符号辞書記憶部2に記憶された符号辞書信号は、「情報」についているキーワードなどの文字列Wを数字Cに変換する対応表をno f DCK個有するコードブック

【数1】

$$DCK[1]=(W[1],C[1])$$

⋮

$$DCK[nofDCK]=(W[nofDCK],C[nofDCK])$$

である。ベクトル生成部1はキーワード数信号no f Ksとno f Ks個のキーワード信号からなるキーワード群信号Ks=(K[1], ⋯, K[nofKs])とを受けキーワード群信号Ksと前記符号辞書信号DCKを用いてベクトル信号Vに変換する。また、スコア計算部3が、ユーザーに提示された「情報」を、必要／不要と評価した結果から計算された肯定メトリック信号MY、否定メトリック信号MNを用いて、ベクトル生成部1で変換された2つのベクトル信号Vの長さで肯定信号SYと否定信号SNとに変換する。肯定メトリック記憶部5は(nofDCK×nofDCK)行列である前記肯定メトリック信号MYを記憶し、否定メトリック記憶部6は(nofDCK×nofDCK)行列である前記否定メトリック信号MNを記憶し、判定パラメータ記憶部8は判定パラメータ信号Cを記憶する部分である。必要性計算部7は前記肯定信号SYと前記否定信号SNを受け前記判別パラメータ記憶部8から判定パラメータ信号Cを

読み出し必要性信号Nと信頼性信号Rを計算する。さらに、未読データ書き込み制御部9が「情報」の本文である情報データDとキーワード数信号nofKsとキーワード群信号Ksと必要性信号Nと信頼性信号Rとを所定の手続きに従って後述する未読データ記憶部10に書き

$$\text{URD}[1]=(\text{N}[1], \text{R}[1], \text{nofKs}[1], \text{Ks}[1], \text{D}[1])$$

⋮

$$\text{URD}[\text{nofURD}]=(\text{N}[\text{nofURD}], \text{R}[\text{nofURD}], \\ \text{nofKs}[\text{nofURD}], \text{Ks}[\text{nofURD}], \\ \text{D}[\text{nofURD}])$$

を記憶する。

【0038】代表情報一覧ユニット53は、未読データ記憶部10に記憶された未読データ相互の類似度を計算し、どの未読データが一まとめにできるかを決定する未読データ圧縮部70、前記未読データ圧縮部70の結果を記憶する圧縮未読データ記憶部81、圧縮未読データを用いて未読データの読み出しを制御する圧縮未読データ読み出し制御部82等からなる。図13に未読データ圧縮部70の構成が示されている。71は類似度計算制御部であり、72は未読データ読み出し部であり、73は類似度ベクトル生成部であり、74はスイッチ、75は参照ベクトル記憶部であり、76は類似度計算部であり、77は閾値記憶部であり、78は比較部であり、79は圧縮未読データ書き込み制御部であり、80はフラグ記憶部であり、81は圧縮未読データ記憶部である。

【0039】インターフェースユニット51は、未読データ出力制御部11が制御信号DOを受け圧縮未読データ読み出し制御部82を通じて未読データ記憶部10から未読データ信号URD[1]を読み出して表示信号DDを出力する。また、その表示信号DDがユーザーにとって必要か否かを示す教師信号Tをユーザーから受け、前記教師信号Tと前記未読データ信号URD[1]のキーワード数信号nofKs[1]とキーワード群信号Ks[1]とを所定の手続きに従って教師データ記憶部13に書き込む。教師データ記憶部13は、最大nofTD個の教師データ信号

【数3】

$$\text{TD}[1]=(\text{T}[1], \text{TnofK}[1], \text{TKs}[1])$$

⋮

$$\text{TD}[\text{nofTD}]=(\text{T}[\text{nofTD}], \text{TnofKs}[\text{nofTD}], \\ \text{TKs}[\text{nofTD}])$$

を記憶する。

【0040】学習ユニット52は、ユーザーから入力された教師信号Tを用いて肯定/否定メトリック信号を修正するメトリック学習を行う部分と、肯定/否定信号から必要性信号を計算するためのパラメータ、判定パラメータ信号を修正する部分からなり、各部分は学習制御部によって制御される。

込む。未読データ記憶部10は、前記「情報」の本文である情報データDと前記キーワード数信号nofKsと前記キーワード群信号Ksと前記必要性信号Nと前記信頼性信号Rとからなる最大nofURD個の未読データ【数2】

【0041】図1において、メトリック学習を行う部分の構成は次の通りである。

【0042】メトリック学習部19において肯定メトリック記憶部5に記憶された肯定メトリック信号MYと前記否定メトリック記憶部6に記憶された否定メトリック信号MNとを修正する。このメトリック学習部19は、教師データ記憶部13から前記教師データTDを読み出し、学習ユニット50のベクトル生成部1と同じ機能である学習用ベクトル生成部20で複数のキーワードをベクトルに変換し、自己相関行列を計算することで、肯定/否定メトリック信号を修正する。

【0043】判定パラメータ信号の学習を行う部分の構成は次の通りである。

【0044】学習用スコア計算部22が、学習用肯定信号計算部221と学習用否定信号計算部222とから構成される。学習用肯定信号計算部221は学習用ベクトル生成部20からの学習用ベクトル信号を受け学習用肯定信号LSYを計算し、学習用否定信号計算部222は学習用ベクトル生成部20からの学習用ベクトル信号を受け学習用否定信号LSNを計算する。判定面学習部21は、学習制御部14からの判定パラメータ学習制御信号PLCを受けて所定の方法で判定パラメータ記憶部8の判定パラメータ信号の書き換えを行う。学習制御部で14は学習開始信号LSを受けてスイッチ16、17、18とメトリック学習部19と学習用ベクトル生成部20と学習用スコア計算部22と学習用否定信号計算部23と判定面学習部21とを制御する。

【0045】以上のように構成された情報フィルタ装置での概略的な処理の流れについて説明する。以下の説明では、既にユーザがどんな「情報」を過去に必要としたかという履歴は学習済みであるものとする。また、単に「情報」と称するものには、「情報」に対応する1つ以上のキーワードが付されているものとする。キーワードは、当該「情報」を構成する各単語の一部あるいは全体であっても良いし、当該「情報」を代表するために特別に付したものであっても良い。

【0046】情報フィルタリングユニット50では、新たな「情報」が入力されると、記憶部2、5、6、8からユーザがどのような「情報」を過去に必要としたかと

いう記録を読みだし、当該ユーザにとっての新たな「情報」の必要性を必要性信号として定量的に評価する。

【0047】次に、その評価された新たな「情報」は、未読データ記憶部10に過去からの未読「情報」を含めて必要性信号が大きい順に並ぶように書き込まれる。

【0048】代表情報一覧ユニット53では、未読データ記憶部10への書き込みが終了すると、未読データ圧縮部70が未読データ記憶部10に記憶された「情報」の相互の類似度を計算し、どの情報が一まとめにできるかという結果を圧縮未読データ信号として圧縮未読データ記憶部81に書き込む。

【0049】インタフェースユニット51では、ユーザーに必要性信号の大きい順に並べて、前記新たな「情報」を含めた未読「情報」、類似の情報がある場合には代表的な情報を提示（例えば、ディスプレイ表示）する。

【0050】この際に、ユーザーに提示された前記新たな「情報」を含めた未読「情報」の1つひとつがユーザーにとって必要か不要かを示す教師信号をユーザーがインタフェースユニット51を介して入力することにより、インタフェースユニット51では、当該教師信号を受け取り、当該「情報」とその教師信号を学習ユニット52に送る。

【0051】なお、このユーザーによる教師信号の入力は、学習ユニット52の学習能力をより高めるために実施するものであり、学習ユニット52の学習能力（ユーザーがどんな「情報」を過去に必要としたかという履歴の学習能力）が既に十分に高ければ行う必要はない。

【0052】次に、学習ユニット52では、前記提示した「情報」とその教師信号を用いて記憶部2、5、6、8の履歴内容を書き換える。

【0053】以上、本発明の情報フィルタ装置は、より高い学習を通じてユーザーに適応し、ユーザーの求める「情報」を優先的に、かつ高い一覧性をもって提示することができる。

【0054】また、学習を行っていない初期状態では、ユーザーがどんな「情報」を必要としているのか学習ユニット52では分からないので、全ての入力される新たな「情報」をインタフェースユニット51でユーザーが提示を受ける毎に上述したユーザーによる教師信号の入力は必要であるが、随時実施する学習を通じてやがてユーザーに適応し、ユーザーの求める「情報」を優先的に提示することができるようになる。

【0055】なお、ユーザーの求める「情報」を優先的に提示するとは、より具体的な使用例で述べれば、ある「情報」データベースの母集団Aを特定のキーワードで検索して「情報」の検索集合Bを得たとしても、当該検索集合Bの「情報」の全てがユーザーにとっては必要であるとは限らないし、またユーザーにとっては「情報」の全てが必要であってもその必要順位は当然存在するこ

とを前提としている。よって、必要から順に不要、あるいはその必要順位に従って、インタフェースユニット51でユーザーに順に提示することが、ユーザーの求める「情報」を優先的に提示することを意味する。

【0056】以下、各ユニットごとの動作内容を図面を用いて具体的に説明する。

【0057】情報フィルタリングユニット50では「情報」の必要性信号を計算する。

【0058】ここで、必要性信号を計算するための基本概念を説明する。

【0059】入力された「情報」にキーワードが添付されている場合を考える。一人のユーザーを考えると、そのユーザーが必要としている「情報」に高い頻度または確率で付いているキーワード集合Aと、不要としている「情報」に高い頻度または確率で付いているキーワード集合Bと、さらにはいずれにもよく付く、または付かないキーワード集合Cとを考えることができる。

【0060】前記キーワード集合Aに属するキーワードには正の数値を、前記キーワード集合Bに属するキーワードには負の値を、前記キーワード集合Cに属するキーワードには値0をそれぞれ割り振る。

【0061】そして、新たに入力された「情報」について1つ以上のキーワードについてそれぞれが前記キーワード集合A、B、Cのどのキーワードグループに属するかを判定し、前記割り振られた値を積算するように構成する。

【0062】このように構成すれば、前記新たに入力された「情報」に付いていた複数のキーワードを、キーワード集合Aに属するキーワードが数多く含まれた「情報」（ユーザーが必要とする可能性の高い情報）に対しては大きな正の値を示し、キーワード集合Bに属するキーワードが数多く付いている「情報」（ユーザーが不要とする可能性の高い情報）に対しては大きな負の値を示す数値に変換することができる。

【0063】こうして、前記数値を用いてユーザーにとっての必要性を予測することができる。本発明では、提示した「情報」とその「情報」に関するユーザーの必要／不要の評価とからキーワードへの値の割り振りを自動的に高い精度の必要性信号の計算を実現し、精度高く必要性の高い順に「情報」を並べ変えることを実現している。

【0064】そのために、「情報」に付けられた複数のキーワードを一つのベクトルに変換し、ユーザーが必要とした場合と不要とした場合について、別々に前記ベクトルの自己相関行列を計算している。ユーザーが必要と答えた「情報」についていたキーワードから作られた自己相関行列MYを用いて、ベクトルVの長さSYを

【数4】

$$SY = \sum_i \sum_j M_{ij} \cdot V_i \cdot V_j$$

と計算する。

【0065】なお、必要と答えた「情報」についていたキーワードから作られた自己相関行列MYを「肯定メトリック信号」、不要と答えた情報についていたキーワードから作られた自己相関行列MNを「否定メトリック信号」と呼び、長さSYを肯定信号と呼んでいる。この長さSYは、ベクトルVの元となった複数のキーワードの中に、ユーザーが必要とする「情報」によく含まれているキーワードが数多く含まれていれば、長さSYは大きな正の値をとり、そうでない場合には0に近い値をとるから、必要性信号を計算する上で有効である。

【0066】情報フィルタ装置の好ましい初期状態の一例は、肯定メトリック信号MYと否定メトリック信号MNとを $(\text{nofDCK} \times \text{nofDCK})$  零行列、未読データ記憶部10の未読データURD[i]の全ての必要性信号N[i] ( $i=1, \dots, \text{nofURD}$ ) を使用するハードウェアが表現可能な最小の値Vmin、教師データ記憶部13の教師データTD[j]の教師信号T[j]を全て-1とした状態である。このような初期状態となっているものとする。

【0067】情報フィルタリングユニット50では、まず、情報データ入力端子100から情報データDが入力され、キーワード数信号入力端子101から情報データに付けられたキーワードの個数を表すキーワード数信号nofKsが入力され、キーワード信号入力端子102から複数のキーワードであるキーワード群信号Ks=(K[1], K[2], ..., K[nofKs])が入力される。

【0068】ベクトル生成部1によって、キーワード群信号Ksが文字列の集まりからベクトル信号Vへと変換される。この変換によって、キーワード群信号の類似性をベクトルの距離として計算できるようになる。

【0069】ベクトル生成部1の動作を図3に示すフローチャートを参照しながら説明する。まず、キーワード数信号nofKsとキーワード群信号Ksを受けると

(図3ステップ(イ))、内部のベクトル信号V=(V[1], V[2], ..., V[nofDic])を(0, 0, ..., 0)に、キーワードカウンタ信号iを1にセットする(同図ステップ(ロ)、(ハ))。次に、辞書カウンタ信号jを0セットした後辞書カウンタ信号jを1だけ増やす(同図ステップ(ニ))。

【0070】次に、内部にnofDCK個の符号辞書信号DCKを有する辞書記憶部2から辞書カウンタjが指定するキーワードと数字からなる符号辞書信号DCK[j]を読み出し、符号辞書信号DCKの文字列部分W[j]とi番目のキーワード信号K[i]とを比較する(同図ステップ(ホ))。両者が等しくない場合には、辞書カウンタjを1だけ増やす(同図ステップ(ヘ))。両者が一致するか、辞書カウンタjの値が辞書記憶部2に格納された符号辞書信号の個数nofDiC

と等しくなるまで図3ステップ(ホ)～(ト)の処理を繰り返す(同図ステップ(ト))。

【0071】キーワード信号K[i]と等しいW[j]が見つかったら、ベクトル信号のj番目の成分V[j]を1にし(同図ステップ(チ))、キーワードカウンタ信号iを1だけ増やす(同図ステップ(リ))。以下、同様の処理をキーワードカウンタ信号iがキーワード数信号nofKsより大きくなるまで実行する(同図ステップ(ヌ))。

【0072】こうして、ベクトル生成部1において、文字列信号からなるキーワード信号の集合体であるキーワード群信号Ksは、0と1でコード化されたnofDCK個の成分を持ったベクトル信号Vに変換される。

【0073】次に、肯定信号計算部31は、キーワード群信号Ksに過去にユーザーの必要とした情報に含まれていたキーワードが数多く含まれる場合に、大きな値となる肯定信号SYを計算する。このために、肯定信号計算部31は、前記ベクトル信号Vを受けて、肯定メトリック記憶部5から肯定メトリック信号MYを読み出し、肯定信号SYを

【数5】

$$SY = \sum_{i=0}^{\text{nofDic}-1} \sum_{j=0}^{\text{nofDic}-1} MY[i][j] \cdot V[i] \cdot V[j]$$

と計算する。

【0074】否定信号計算部32は、キーワード群信号Ksに過去にユーザーの不要とした情報に含まれていたキーワードが数多く含まれる場合に、大きな値となる否定信号SNを計算する。このために、否定信号計算部32は、否定メトリック記憶部6から否定メトリック信号MNを読み出し、否定信号SNを

【数6】

$$SN = \sum_{i=0}^{\text{nofDic}-1} \sum_{j=0}^{\text{nofDic}-1} MN[i][j] \cdot V[i] \cdot V[j]$$

と計算する。

【0075】肯定メトリック信号MYと否定メトリック信号MNは、後述するようにキーワード群信号Ksとユーザーの応答に基づいて決められる。

【0076】このように計算された肯定信号SYと否定信号SNを用いて、図9に示したように縦軸に肯定信号SYをとり横軸に否定信号SNをとった2次元空間上の1点に、情報データDを対応させることができる。この2次元空間における情報データDの分布は、ユーザーが必要とするもの(○で表示)は主に左上部に分布し、ユーザーが不要とするもの(×で表示)は主に右下部に分布するようになる。

【0077】したがって、図10に示したように適切な係数Cを定めることにより、ユーザーが必要とする情報データDと不要な情報データDとを分離できる。

【0078】さらに、以下に述べるこの係数Cを用いて計算される必要性信号Nは、上述の2次元空間で左上にある程、すなわち、必要性の高いと予測される情報データDほど大きな値となる。したがって、必要性信号Nの大きい順に情報データDを並べて提示すれば、ユーザーは必要な情報を効率よく手に入れることができる。必要性信号Nと直交する方向の信頼性信号Rは、大まかにはキーワード群信号Ksに含まれていたキーワードのうちのどれくらいのキーワード信号が辞書に含まれていたかを示す信号である。したがって、この信頼性信号Rの大きさは、情報フィルタが計算した必要性信号Nがどれだけ信頼できるのかを示す。

【0079】次に、必要性計算部7は、前記肯定信号計算部31から出力される前記肯定信号SYと前記否定信号計算部32から出力される前記否定信号SNとを受け、判定パラメータ記憶部8から判定パラメータ信号Cを読み出し、過去必要であった情報についていたキーワードが多数あり、不要であった情報についていたキーワードがほとんどない時に大きな値となる必要性信号Nを  $N = SY - C \cdot SN$  と計算し、信頼性信号Rを  $R = C \cdot SY + SN$  と計算する。

【0080】未読データ書き込み制御部9の動作を、図4に示したフローチャートを参照しながら説明する。まず、それぞれの入力端子から前記情報データDと前記キーワード数信号nofKsと前記キーワード群信号Ksを受け、必要性計算部7から前記必要性信号Nと前記信頼性信号Rとを受け、未読データ部指示端子110から出力する未読データ処理信号WIを0から1に変える（図4ステップ（イ））。次に、 $i = 1$  とし（同図ステップ（ロ））、未読データ記憶部10に記憶された未読データURD[i]の必要性信号N[i]（ $i = 1, \dots, \text{nof URD}$ ）を順次読み出し、前記必要性信号Nと比較し（同図ステップ（ハ））、前記必要性信号Nが未読データURD[i]の必要性信号N[i]より大きくなる（ $N \geq N[i]$ ）最初の未読データの番号i1を検出する（同図ステップ（ニ）（ホ））。

【0081】i1番目以降の未読データを  $URD[i+1] = URD[i] \quad i = i1, \dots, \text{nof URD}$  と置き換え（同図ステップ（ヘ）～（リ））、その後、i1番目の未読データURD[i1]を  $N[i1] = N$   
 $R[i1] = R$   
 $\text{nof Ks}[i1] = \text{nof Ks}$   
 $Ks[i1] = Ks$   
 $D[i1] = D$  と前記必要性信号N等で置き換える（同図ステップ（ヌ））。この置き換えが終了すると、未読データ部指

示端子110から出力する未読データ部指示信号WIを0に戻し（同図ステップ（ル））、処理を終了する。

【0082】次に、代表情報一覧ユニット53では、未読データ記憶部70に端子170から類似度計算開始信号が入力されると、類似度計算制御部71は、内部の情報カウンタnと圧縮情報カウンタaをそれぞれ1にセットする（図14ステップ（イ））。

【0083】前記情報カウンタのカウント出力nと圧縮情報カウンタのカウント出力aの値を圧縮未読データ書き込み制御部79に送る。圧縮未読データ書き込み制御部79は、類似情報リスト信号AD[a][1]に前記情報カウンタnの値を代入し、フラグA[a]に類似の情報がないことを示す値1を代入し、比較回数カウンタのカウント出力lに1を代入し、比較対象情報番号差カウンタΔmに0、比較情報番号カウンタのカウント出力mに前記情報カウンタの値に1を加えたものを代入する（同図ステップ（ロ））。

【0084】類似度計算制御部71は、情報カウンタnの値を未読データ読み出し部72に送り、スイッチ74を参照ベクトル記憶部75が接続するように切り替える。未読データ読み出し部72は、未読データ記憶部10から、未読データURD[n]を読み出し（同図ステップ（ハ））、類似度ベクトル生成部73に送る。

【0085】類似度ベクトル生成部73は、前記ベクトル生成部1と同様にして類似度ベクトル信号を生成し（同図ステップ（ニ））、参照ベクトル記憶部75に送る。参照ベクトル記憶部75は、前記類似度ベクトル信号を参照ベクトル信号Vnとして記憶する。類似度計算部76は、参照ベクトル記憶75から参照ベクトル信号Vnを読み出し、

【数7】

$$\sum Vn[i]$$

を計算する（同図ステップ（ホ））。

【0086】類似度計算制御部71は、スイッチ74を類似度計算部76につながるように切り替え、比較情報番号カウンタのカウント出力mの値を圧縮データ書き込み制御部79から読み出し、未読データ読み出し部72に送る。

【0087】未読データ読み出し部72は、未読データ記憶部10から未読データURD[m]を読み出し（同図ステップ（ヘ））、その未読データURD[m]を類似度ベクトル生成部73に送る。

【0088】類似度ベクトル生成部73は、前記ベクトル生成部1と同様にして類似度ベクトル信号を生成し（同図ステップ（ト））、比較ベクトル信号Vmとして類似度計算部76に送る。

【0089】類似度計算部76は、比較ベクトル信号Vmから、

【数8】



$$\sum_i Vm[i]$$

と

【数9】

$$\sum_i Vn[i] \cdot Vm[i]$$

とを計算する(同図ステップ(ホ)、(リ))。類似度計算部76は、

【数10】

$$\sum_i Vm[i]$$

が

【数11】

$$\sum_i Vn[i]$$

より大きい場合には、類似度Chを

【数12】

$$Ch = \frac{\sum_i Vn[i] \cdot Vm[i]}{\sum_i Vn[i]}$$

とし、その他の場合には、

【数13】

$$Ch = \frac{\sum_i Vn[i] \cdot Vm[i]}{\sum_i Vm[i]}$$

とする。この類似度Chは、未読データURD[n]につけられたキーワードのすべてが、未読データURD[m]につけられたキーワードに含まれる場合(または、未読データURD[m]につけられたキーワードのすべてが、未読データURD[n]につけられたキーワードに含まれる場合)に1となり、その他の場合には1より小さくなるという性質を持っている。

【0090】以下に、この類似度の性質について、従来よく使われているベクトルの角度と比較して述べる。

【0091】よく用いられるベクトルの角度は、2つのベクトルが完全に等しい場合しか0にならず、つけられたキーワードに違いがあるか否かを表現する上では有効であるが、2つのキーワード集合の間に一方が他方を包含する場合(図17(a))と2つの集合が重ならない部分を持つ場合(図17(b))とを区別できないという性質を持つ。このため、キーワードの包含関係を表現する上では適切でない。

【0092】実際、ある事件や事柄について、詳しい情報とおおまかな情報とがある場合を考えると、詳しい情報につけられたキーワードの集合がおおまかな情報につけられたキーワードを含むということはしばしば見られる。

【0093】したがって、情報の類似度考える場合、上

の類似度Chは、情報の類似度を評価するために、従来よく使われているベクトルの角度より有効である。

【0094】さて、比較部78は、こうして得られた類似度Chをあらかじめ閾値記憶部77に書き込んでいた閾値 $\theta_{ch}$ と比較する(図15ステップ(ワ))。

【0095】類似度Chが閾値 $\theta_{ch}$ より大きい場合、圧縮未読データ書き込み制御部79は、2つ未読データURD[n]、URD[m]は同じ内容の情報とし、未読データURD[m]を未読データURD[n]の従属情報とする。このために、フラグ信号flag[m]を従属情報であることをあらわす値1にし、圧縮情報カウンタA[a]の値を1だけ増やし、類似情報リスト信号AD[a][A[a]]に比較情報番号カウンタmの値を書き込む(同図ステップ(カ))。

【0096】類似度Chが閾値 $\theta_{ch}$ より小さい場合、2つ未読データURD[n]、URD[m]は異なる内容の情報とし、未読データURD[m]を未読データURD[n]の独立情報とする。このために、フラグ信号flag[m]を独立情報であることをあらわす値0にする(同図ステップ(コ))。

【0097】次に、比較対象情報番号差カウンタ $\Delta m$ に1を加える(同図ステップ(タ))。

【0098】比較情報番号カウンタmに比較対象情報番号差カウンタの値 $\Delta m$ を加え、未読データ数nofURDと比較する(同図ステップ(レ))。(m+ $\Delta m$ )がnofURDであればまだ比較すべき未読データがあることになる。そこで、未読データURD[m+ $\Delta m$ ]が独立データであるか否かを調べる(同図ステップ(ソ))。つまり、すでに従属情報となった情報については、重ねて従属情報とする処理は行わないということである。

【0099】独立情報でない場合には、比較対象情報番号差カウンタ $\Delta m$ に1を加え(同図ステップ(タ))、独立情報が見つかるか、未読データがなくなるまで同様の処理を行う。独立情報が見つかったと、比較回数カウンタ1の値を1だけ増やし、比較情報番号カウンタmに(m+ $\Delta m$ )を代入する(同図ステップ(ツ))。比較回数カウンタ1の値があらかじめ設定しておいた最大比較回数lmax以下ならば、新たに未読データURD[m]を読み出して同様の処理を行う(同図ステップ(ネ))。

【0100】こうして、未読データURD[n]を代表の情報とする従属情報のリストが、類似情報リスト信号AD[a][A[a]]として作られる。

【0101】比較回数カウンタ1の値があらかじめ設定しておいた最大比較回数lmaxより大きくなる(同図ステップ(ネ))と、新たな未読データを代表の情報とする従属情報を探す手続きが始まる。そのためには、上と同様に、従属情報として処理されていない未読データを探す(同図ステップ(ナ)、(ラ)、(ム))、

(ウ))。

【0102】新たに、従属情報として処理されていない未読データが見つかったと、情報カウンタ $n$ の値を $(n + \Delta n)$ 、圧縮情報カウンタ $a$ の値を1だけ増やし、ステップ(ロ)に戻り上記の処理を行う。

【0103】この処理を、情報カウンタ $n$ が未読データの数 $\text{nof URD}$ 以上となるまで繰り返す。

【0104】この処理で得られる類似情報リスト信号 $AD$ は、図16のようなものとなる。圧縮未読データ書き込み部79は、類似度情報信号 $AD$ と圧縮情報カウンタの値、すなわち独立情報の数とを圧縮未読データおよび圧縮未読データ数 $\text{nof a}$ として、圧縮未読データ記憶部81に書き込む。

【0105】次に、圧縮未読データ読み出し制御部と、未読データ $UDR$ を読みだし、ユーザーの応答(教師信号 $T$ )を付加して教師データ信号 $TD$ をつくるインターフェースユニット51について説明する。

【0106】次に、インターフェースユニット51の動作を図5に示したフローチャートを参照しながら説明する。

【0107】データ読み出し開始信号入力端子103から、データ読み出し開始信号 $DO$ が入力される(図5ステップ(イ))。未読データ出力制御部11は、圧縮未読データ制御部82を通じて、前記未読データ記憶部10から1番目の独立情報である未読データ $URD[AD[1][1]]$ を読み出し(同図ステップ(ロ))、未読データの必要性信号 $N[1]$ が最小値 $V_{min}$ より大きい場合には、未読データ信号 $URD[1]$ の情報信号 $D[1]$ を表示情報信号 $DD$ としてデータ表示端子104に出力し、待機する(同図ステップ(ハ))、

(ニ))。未読データの必要性信号 $N[1]$ が最小値 $V_{min}$ に等しい場合には、表示情報信号 $DD$ を「データなし」としてデータ表示端子104に出力し、待機する(同図ステップ(ホ))。

【0108】ユーザー(図示せず)は、データ表示装置(図示せず)に表示された表示情報信号 $DD$ を見て、それが必要な情報である場合には教師信号 $T=1$ 、必要でない場合には教師信号 $T=0$ 、処理を終了する場合には教師信号 $T=-1$ として、教師信号入力端子105に返す(同図ステップ(ヘ))。教師信号 $T=-1$ の場合、処理を終了し、教師信号 $T \neq -1$ の場合には(同図ステップ(ト))、未読データ出力制御部11は、教師データ記憶部13の(数2)で表わされる教師データを $TD[i] = TD[i-1]$ ,  $i=2, \dots, \text{nof TD}$

と置き換え(同図ステップ(ル))、1番目の教師データ $TD[1]$ を前記教師信号 $T$ と前記未読データのキーワード数信号 $\text{nof Ks}[1]$ とキーワード群信号 $Ks[1]$ とを用いて

$T[1] = T$

$\text{Tnof Ks}[1] = \text{nof Ks}[1]$

$TKs[1] = Ks[1]$

と置き換え(同図ステップ(ル))、(ヲ))、前記未読データ記憶部10の未読データ $URD$ を

$URD[AD[a][b]] = URD[AD[a+1][b]]$ ,  $a=1, \dots, (\text{nof a}-1)$ ,  $b=1, 2, \dots$

とし(同図ステップ(ワ))、(カ))、 $\text{nof a}$ 番目の独立した未読データ $URD[AD[\text{nof a}][1]]$ の必要性信号を

$N[\text{nof a}] = (\text{最小値 } V_{min})$

とする(同図ステップ(ヨ))、(タ))、(レ))。

【0109】次に、学習ユニット52の動作について図6～図8に示したフローチャートを参照しながら説明する。

【0110】図6に学習制御部14の動作の概略を示すフローチャート示し、詳しく説明する。

【0111】図6において、まず、学習開始信号入力端子106から学習開始信号 $LS$ が入力され、学習制御部指示信号出力端子107から出力される学習制御部指示信号 $LI$ を0から1に変え(図6ステップ(イ))、処理中を示す。次に、スイッチ16とスイッチ17とスイッチ18とをメトリック学習部19と学習用ベクトル生成部20が接続する様に切り替える(同図ステップ(ロ))。

【0112】次に、図7のステップ(ハ)に対応するメトリック学習部19を動作し(同図ステップ(ハ))、判定面学習部21を動作させた後(同図ステップ(ニ))、 $LI$ を0として(同図ステップ(ホ))、処理を終了する。

【0113】次に、メトリック学習部19がユーザーの応答(教師信号 $T$ )とキーワード群信号 $Ks$ とを用いて、肯定/否定メトリック信号を修正する動作を図7を用いて説明する。

【0114】図7は、メトリック学習部19の動作のフローチャートであり、同図において、前記学習制御部14からメトリック学習制御信号 $MLC$ を受けた(図7ステップ(イ))メトリック学習部19は、肯定メトリック記憶部5から肯定メトリック信号 $MY$ を、否定メトリック記憶部6から否定メトリック信号 $MN$ をそれぞれ読み出す。

【0115】次に、メトリック学習部19は、教師データカウンタ $c$ の値を1にする(同図ステップ(ロ))。次に教師データ記憶部13から $c$ 番目の教師データ信号 $TD[c]$ を読み出し(同図ステップ(ハ))、教師データ $TD[c]$ の教師信号 $T[c]$ を調べる。前記教師信号 $T[c]$ が-1でない場合( $T \neq -1$ )には(同図ステップ(ニ))、教師データ $TD[c]$ のキーワード数信号 $\text{Tnof Ks}[c]$ とキーワード群信号 $TKs[c]$ とを出力する(同図ステップ(ホ))。前記教師

データTD[c]のキーワード数信号TnofKs[c]とキーワード群TKs[c]とを受けた学習用ベクトル生成部20は、前述の情報フィルタリングユニット50のベクトル生成部1と同様の動作を行い、学習用ベクトル信号LVを出力する(同図ステップ(ヘ))。メトリック学習部19は、前記学習用ベクトル信号LVを受け、前記教師データTD[c]の教師信号T[c]がT=1である場合には(同図ステップ(ト))、肯定メトリック信号MYを

$$MY[i][j] = MY[i][j] + LV[i] \cdot LV[j]$$

(ここで、 $i, j = 1 \sim \text{nofDiC}$ )と修正する(同図ステップ(チ))。

【0116】この処理により、肯定メトリック信号MYは、ユーザーが必要とした情報データDについていたキーワード信号(複数)に対して大きな値を持つようになる。その結果、前述の肯定信号SYが、ユーザーが必要とする情報データDに対して大きくなるようになる。否定メトリック信号MNも以下のように同様の処理がなされる。

【0117】前記教師データTD[c]の教師信号T[c]がT=0である場合には、否定メトリック信号MNを

$$MN[i][j] = MN[i][j] + LV[i] \cdot LV[j]$$

(ここで、 $i, j = 1 \sim \text{nofDiC}$ )と修正する(同図ステップ(リ))。

【0118】教師データカウンタの値を

$$c = c + 1$$

と1だけ増やす(同図ステップ(ヌ))。

【0119】以下、メトリック学習部19は、同様の動作を、教師データTD[c]の教師信号T[c]がT[c]=-1になるかまたは $c = \text{nofTD}$ となるまで繰り返す。T[c]=-1または $c = \text{nofTD}$ となると(同図ステップ(ヲ))、メトリック学習の処理を終了し、メトリック学習制御信号MLCを学習制御部14に送る。

【0120】学習制御部14は、メトリック学習部19からのメトリック学習制御信号MLCを受けて、スイッチ16を学習用ベクトル生成部20とスコア計算部22とが接続するように切り替え、スイッチ17とスイッチ18を学習用ベクトル生成部20と判定面学習部21とが接続するように切り替える。学習制御部14は、判定面学習制御信号PLCを判定面学習部21に送る。

【0121】次に、判定面学習部21について、図8を用いて詳しく説明する。

【0122】判定面学習部21は、図10に示したように、肯定信号SYと否定信号SNを用いて2次元空間上に表現されたユーザーが必要とする情報データDとユーザーが不要とする情報データDをもっともよく分離す

る係数Cを求める。

【0123】このための処理を、図8に示したフローチャートに従って詳しく説明する。

【0124】まず、前記判定面学習制御信号PLCを受けて(図8ステップ(イ))、教師データカウンタcの値を1にする(同図ステップ(ロ))。教師データ記憶部13からc番目の教師データ信号TD[c]を読み出し(同図ステップ(ハ))、教師データTD[c]の教師信号T[c]を調べる(同図ステップ(ニ))。前記教師信号T[c]が-1でない場合( $T \neq -1$ )には、教師データTD[c]のキーワード数信号TnofKs[c]とキーワード群信号TKs[c]とを出力する(同図ステップ(ホ))。前記教師データTD[c]のキーワード数信号TnofKs[c]とキーワード群TKs[c]とを受けた学習用ベクトル生成部20は、前述した情報フィルタリングユニット50のベクトル生成部1と同様の動作を行い、学習用ベクトル信号LVを出力する。

【0125】学習用スコア計算部22は、前述した情報フィルタリングユニット50のスコア計算部3と同様の動作を行い、学習用肯定信号LSY[c]と学習用否定信号LSN[c]とを出力し、判定面学習部21がそれを受ける(同図ステップ(ヘ))。前記学習用肯定信号LSY[c]と前記学習用否定信号LSN[c]と教師データTD[c]の教師信号T[c]と判定面学習用信号TC[c] = (T[c], LSN[c], LSY[c])を内部の記憶素子に記憶する(同図ステップ(ト))。そして、教師データカウンタの値を

$$c = c + 1$$

と1だけ増やす(同図ステップ(チ))。

【0126】以下、判定面学習部21は、同様の動作を、教師データTD[c]の教師信号T[c]がT[c]=-1になるかまたは $c = \text{nofTD} + 1$ となるまで繰り返す(同図ステップ(リ))。T[c]=-1または $c = \text{nofTD}$ となると、学習用肯定信号LSY[c]計算等の処理を終了する。

【0127】次に、判定面学習部21は、内部の記憶素子に記憶された判定面学習用信号TC[c] ( $c = 1, \dots$ )は、横軸をLSN[c]、縦軸をLSY[c]とし、T[c]=1を○、T[c]=0を×で示すと、図9に示すような分布となる。これらのうち、教師信号T[c]=1であるものと前記教師信号T[c]=0であるものが、図10に示したように最もよく分離できる判定パラメータCを、山登り法によって計算する(同図ステップ(ヌ))。次に前記判定パラメータCを判定パラメータ記憶部8に書き込み、学習制御部14に判定面学習制御信号PLCを送り(同図ステップ(ル))、処理を終了する。学習制御部14は、判定面学習部21から判定面学習制御信号PLCを受け、学習制御部指示信号を待機中を示す値にし、処理を終了する。

【0128】図10に示したように、上述の2つのメトリック信号を用いてキーワード群信号を肯定信号SYと否定信号SNとで表される2次元空間上で、ユーザーが必要とする情報は主に左上に、不要な情報は右下に分布するようになる。したがって、上記のように適切な係数Cを用いて必要性信号を $N = SY - C \cdot SN$ とすれば、必要性信号は、ユーザーが必要とする情報に対して大きな値をとるようになる。

【0129】なお、判定パラメータCの計算方法として、ここでは、山登り法を採用したが、判定面と学習用必要性信号LNと学習用信頼性信号LRとの距離に基づいて構成されるコスト関数

【数14】

$$COST = \sum_c (2 \cdot T[c] - 1) (LSN[c] - C \cdot LSY[c])$$

を最大にする判定面パラメータCをニュートン法、挟み撃ち法などで求める方法であってもよい。

【0130】また、肯定メトリック信号MYと否定メトリック信号MNの学習を忘却の効果を入れた

$$MY[i][j] = \alpha \cdot MY[i][j] + LV[i] \cdot LV[j]$$

$$MN[i][j] = \beta \cdot MN[i][j] + LV[i] \cdot LV[j]$$

を用いてもよい結果が得られる。(ここで、 $\alpha$ と $\beta$ とは、1より小さい正の数)さらに、文献「情報処理学会技術報告、自然言語処理101-8(1994.5.27)」などに記載された文書からキーワード群信号とキーワード数信号を生成するキーワード生成部を付加する構成をとれば、キーワードが与えられていない情報に対しても適用できる情報フィルタ装置を構成することができる。

【0131】タイトルがつけられた情報については、タイトルを構成する単語をもってキーワードとし、キーワード数信号とキーワード群信号を生成してもよい。

【0132】加えて、キーワード信号は、国際特許分類番号など分類記号を含むようにしても、本発明の構成を変更する必要はなく、よい結果をえることができる。

【0133】また、本発明の実施の形態1では、未読データURDを1つずつ提示する場合について示したが、表示装置(図示せず)の大きさによっては、複数の未読データURDを同時に表示し、ユーザが複数表示されたなどの未読データに対して応答したのが正しく情報フィルタ装置に伝えられるような構成を取ることは容易である。

【0134】また、本発明の実施の形態1では、代表情報を含む独立情報だけを表示するように構成したが、ユーザーの要求によって、従属情報の数又は有無を合わせて表示させるようにしてもよい。また、必要に応じて、独立情報の一覧性を損なわない範囲で、従属情報を小さな文字、画像で表示してもよい。

【0135】実施の形態1の情報フィルタ装置によれば、情報相互の類似度を計算して類似情報をひとまとめにするとともに、代表情報及び独立情報を一覧表示するようにしたので、ユーザーに必要性の順に情報を提示する場合に、同じ内容の情報が何度も繰り返し提示されるといった不具合を防止できる。さらに、本発明の主たる目的ではないが、本発明を単純な「類似情報を一まとめにする装置」として用いることも可能である。

【0136】実施の形態1の情報フィルタ装置によれば、ユーザーの応答とキーワードとの関係をキーワードの同時出現に注目した肯定メトリック信号MY、否定メトリック信号に反映させ、この2つのメトリック信号を用いてキーワード群信号を肯定信号SYと否定信号SNとに変換することで、キーワードという記号情報を距離の定義された空間に射影したものである。これによって、キーワード群の遠近を距離というアナログ尺度で評価することができるようになる。これを利用することにより、従来の技術では必要か不要かの二者択一的な判定しかできなかった必要性の評価が、ユーザーの必要性の順番に並べるといったことが可能になる。

【0137】実施の形態1の情報フィルタ装置によれば、ユーザーからの教師信号に基づいた学習によって、ユーザーの必要とする情報に対しては、必要性信号が大きな値を取るようになり、その結果、表示装置等のインターフェースユニットには、ユーザーにとって必要性が高い情報が優先的に、かつ類似の情報は代表的なものが一つ表示されるようになる。

【0138】(実施の形態2)以下、本発明の第2の実施の形態について、図面を参照しながら説明する。実施の形態2は、実施の形態1の構成に辞書学習部を付加し、辞書記憶部2に記憶された符号辞書信号DCKがユーザに適応するように更新し、かつ肯定メトリック信号MYと否定メトリック信号MNを単純な頻度分布に対応するキーワードの自己相関行列から、情報が必要/不要の出現するキーワードの確率分布を考慮したものへと改良したものである。

【0139】図11に本発明の実施の形態2の情報フィルタ装置のブロック結線図を示すが、前述した本発明の実施の形態1の情報フィルタ装置のブロック結線図と異なる構成について詳細に説明する。

【0140】図11において、23は学習制御部14からの辞書学習信号DLCを受け辞書記憶部2の符号辞書信号DCKを更新する辞書学習部、24は文字列Wと数字Cがキーワード群信号Ksに含まれていたときにユーザが情報データDを必要と解答した回数を示す肯定回数PYと文字列Wがキーワード群信号Ksに含まれていたときにユーザが情報データDが不要と解答した回数を示す否定回数PNとからなる表をnofDCK個有する適応符号辞書信号

【数15】

```

FDCK[1]=(W[1],C[1],PY[1],PN[1])
:
FDCK[nofFDCK]=(W[nofFDCK],C[nofFDCK],
PY[nofFDCK],PN[nofFDCK])

```

を記憶した適応符号辞書信号記憶部、25はユーザが必要と答えた回数を示す全肯定回数信号NYと不要と答えた回数を示す全否定回数信号NNを記憶する回数記憶部、26は肯定メトリック更新用の1次肯定メトリック信号MY1を記憶する1次肯定メトリック記憶部、27は否定メトリック更新用の1次否定メトリック信号MN1を記憶する1次否定メトリック記憶部、28は前記肯定回数信号と前記否定回数信号と前記1次肯定メトリック信号MY1と前記1次否定メトリック信号MN1とから改良された肯定メトリック信号MYと否定メトリック信号MNを計算してそれぞれを肯定メトリック記憶部5と否定メトリック記憶部6に書き込むKDメトリック学習部である。

【0141】以上のように構成された情報フィルタ装置について、図面を用いて動作を説明する。ただし、動作が実施の形態1と同様の箇所は説明を省略する。

【0142】情報フィルタ装置の好ましい初期状態の一例は、肯定メトリック信号MYと否定メトリック信号MNとを(nofDCK×nofDCK)零行列、未読データ記憶部10の未読データURD[i]の全ての必要性信号N[i] (i=1, ..., nofURD)を使用するハードウェアが表現可能な最小の値Vmin、教師データ記憶部13の教師データTD[j]の教師信号T[j]を全て-1、適応符号辞書信号の文字列Wを全てブランク、数字Cを符号辞書信号FDCKの上から順に1、2、...、nofFDCK、肯定回数PYと否定回数PNを0、適応符号辞書に対応して、符号辞書の文字列も全てブランクとした状態である。

【0143】まず、情報フィルタリングユニット50の動作を説明する。

【0144】上述の初期状態の場合、実施の形態1に記載した通りの動作を情報フィルタリングユニット50は行い、入力されたキーワード数信号nofKs、キーワード群信号Ks、情報データDから必要性信号N、信頼性信号Rをとともに0と計算し、未読データ記憶部10に格納する。次に、インタフェースユニット51は、実施の形態1と同じ動作を行い、ユーザの応答が付いた教師データTDを教師データ記憶部13に送る。

【0145】未読データ圧縮部70の動作も、実施の形態1と同様である。

【0146】学習ユニット52の動作は、まず、学習開始信号入力端子106から学習開始信号LSが入力される。学習制御部14は、前記学習開始信号LSを受けて、学習制御部指示信号出力端子107から出力される

学習制御部指示信号LIを0から1に変え、処理中を示す。更に、辞書学習信号DLCを辞書学習部23に送る。

【0147】以下、図12に示したフローチャートを参照しながら辞書学習部23の動作を説明する。まず、辞書学習信号DLCを受けて(図12ステップ(イ))、適応符号辞書記憶部24から適応符号辞書FDCKを最大nofFDCKtmp個の適応符号信号を記憶できる適応符号信号バッファに読み込み、回数記憶部25から全肯定回数信号NYと全否定回数信号NNとを、1次肯定メトリック記憶部26から1次肯定メトリック信号MY1を、1次否定メトリック信号記憶部27から1次否定メトリック信号MN1を読み出す(同図ステップ(ロ))。次に内部の教師データカウンタcの値を1にし(同図ステップ(ハ))、教師信号記憶部13から教師データTD[c]を読み出し(同図ステップ(ニ))、その教師信号T[c]が-1であるか否かを判定する(同図ステップ(ホ))。

【0148】T[c]≠-1の場合、以下の処理を行う。まず、内部のキーワード数カウンタiの値を1にセットし(同図ステップ(ヘ))、適応符号辞書カウンタjの値を1にセットする(同図ステップ(ト))。次に、前記文字列W[j]がブランクであるかないかを判定し(同図ステップ(チ))、ブランクである場合には、前記文字列W[j]を前記キーワード信号TK[i]で置き換える(同図ステップ(リ))。ブランクでない場合には、教師データTD[c]のi番目のキーワード信号TK[i]とj番目の適応符号辞書信号FDCK[j]の文字列W[j]とを比較する(同図ステップ(ヌ))。

【0149】前記文字列W[j]がブランクの場合、または、ブランクでなくかつ前記キーワード信号TK[i]と前記文字列W[j]が一致した場合、T[c]の値に応じて以下の処理を行う。T[c]=1の場合(同図ステップ(ル))、全肯定信号NYに1を加え(同図ステップ(ヲ))、適応符号辞書信号FDCK[j]の肯定回数PY[j]に1を加える(同図ステップ(ワ))。T[c]≠1、これはT[c]=0の場合であるが、全否定信号NNに1を加え(同図ステップ(カ))、適応符号辞書信号FDCK[j]の否定回数PN[j]に1を加える(同図ステップ(ヨ))。

【0150】前記W[j]がブランクでなくかつ前記キーワード信号TK[i]と前記文字列W[j]が一致しない場合、適応符号辞書カウンタjの値を1増やす(同

図ステップ(タ))。適応符号辞書カウンタjの値が適応符号辞書信号バッファに記憶できる適応符号信号の数に1を加えた値n of FDCKtmp+1と比較する(同図ステップ(レ))。適応符号辞書カウンタjの値が、n of FDCKtmp+1以下の場合、文字列W[j]がブランクかどうかの判定に戻る。

【0151】それ以外の場合は、前記キーワードカウンタiの値を1だけ増やす(同図ステップ(ソ))。

【0152】前記キーワードカウンタiの値が、前記教師データTD[c]のキーワード数信号T n of Ksに1を加えた値T n of Ks+1と比較して小さい場合(同図ステップ(ツ))、辞書カウンタjを1にセットし、同様の処理を行う。それ以外の場合、教師データカウンタcの値を1だけ増やす(同図ステップ(ネ))。教師データカウンタcの値と教師データ数n of TDに1を加えた値n of TD+1とを比較し(同図ステップ(ナ))、教師データカウンタcの値が小さい場合、次の教師データTD[c]を読み出し同様の処理を行う。

【0153】以上の処理が、全ての教師データTDに対して行われる。

【0154】次に、辞書学習部23は、各々の適応符号辞書信号FDCK[j]に対し、キーワードコスト信号KDを計算する。このキーワードコスト信号は、文字列W[j]がキーワードとして有効であるか否かを判断するために用いられる量である。

【0155】ところで、ユーザの不要な情報データDが出現する確率

$$NN/(NY+NN)$$

と比較して、文字列W[j]が付いている情報データDがユーザにとって不要である場合の確率

$$PN[j]/(PY[j]+PN[j])$$

が大きく異なる場合に、大きくなるようものであれば、文字列W[j]は、情報データDがユーザにとって不要と判定する上で有効である。同様に、ユーザの必要な情報データDが出現する確率

$$NY/(NY+NN)$$

と比較して、文字列W[j]が付いている情報データDがユーザにとって必要である場合の確率

$$PY[j]/(PY[j]+PN[j])$$

が大きく異なる場合に、大きくなるようものであれば、文字列W[j]は、情報データDがユーザにとって必要と判定する上で有効である。

【0156】キーワードコスト信号KDは、この性質を反映している量で有ればなんでもよいが、好ましい例の一つとして、カルバックダイバージェンスと呼ばれる

【数16】

$$NY/(NY+NN) \cdot \log((PY[j])/(PY[j]+PN[j])) + NN/(NY+NN) \cdot \log((PN[j])/(PY[j]+PN[j]))$$

が考えられる。しかし、これは、そのままでは、本情報

フィルタ装置の初期状態など、全肯定回数信号NY、全否定回数信号NN、肯定回数PY[j]、否定回数PN[j]が0のときには、log()の計算ができない、

$$PY[j]+PN[j] \approx 1$$

を満たす適応符号辞書信号FDCK[j]のキーワードコスト信号を過大評価する等不適切な場合がある。これを回避する好ましい実施の形態の一つは、キーワードコスト信号を

【数17】

KD[j]

$$\begin{aligned} &= \tanh((PY[j]+PN[j])/PC) \cdot \\ &\quad \tanh\{NY/(NY+NN) \\ &\quad \cdot \log((PY[j]+\epsilon)/(PY[j]+PN[j]+2\epsilon)) \\ &\quad + NN/(NY+NN) \\ &\quad \cdot \log((PN[j]+\epsilon)/(PY[j]+PN[j]+2\epsilon)) \} \end{aligned}$$

とするものである。ここで、 $\epsilon$ は0でのわり算、log0を避けるための小さな正の値を持つパラメータである。パラメータPCは、3-20程度の値とするとよい。

【0157】次に、適応符号辞書信号FDCK[j]の文字列W[j]と肯定回数PY[j]と否定回数PN[j]とをキーワードコスト信号KDの大きい順に並べ替える(同図ステップ(ラ))。このとき、適応符号辞書FDCK[j]の数字C[j]には、最初の並び順が残っている。これを利用して、1次肯定メトリック信号MY1とC[j]から、C[i]、C[j]の値がともに符号辞書DCKの数n of DCKより小さい場合、 $M[i][j] = MY1[C[i]][C[j]]$ 、 $i, j = 1, \dots, n \text{ of DCK}$

その他の場合は、 $i = j$ の場合は、

$$M[i][i] = PY[C[i]], \quad i = 1, \dots, n \text{ of DCK}$$

$i \neq j$ の場合は、

$$M[i][j] = 0, \quad i, j = 1, \dots, n \text{ of DCK}$$

とした上で、

$$MY1[i][j] = M[i][j], \quad i, j = 1, \dots, n \text{ of DCK}$$

と1次肯定メトリック信号MY1の置き換えを行う。1次否定メトリック信号MN1に対しても、同様の置き換えを行う(同図ステップ(ム))。

【0158】そして、適応符号辞書信号バッファ内の適応符号辞書FDCK[j]の数字C[j]を $C[j] = j, j = 1, \dots, n \text{ of FDCKtmp}$ と置き換える。

【0159】以上の処理を終えると、辞書学習部23は、適応符号辞書バッファ内の適応符号辞書FDCKの上位n of DCK個の文字列W[j]と数字C[j]を辞書記憶部2に書き込み、適応符号辞書バッファ内の適応符号辞書信号FDCK[j]の上位n of FDCK個を適応符号辞書記憶部24に書き込み、全肯定回数信号NYと全否定回数信号NNを回数記憶部25に書き込み、1

次肯定メトリック信号MY1を1次肯定メトリック信号記憶部26に1次否定メトリック信号MN1を1次否定メトリック信号記憶部27に書き込む(同図ステップ(ウ))。

【0160】最後に、辞書学習信号DCLを学習制御部14に戻して(同図ステップ(ヒ))、処理を終了する。

【0161】次に、前記学習制御部14は、スイッチ16とスイッチ17とスイッチ18とをメトリック学習部19と学習用ベクトル生成部20が接続する様に切り替える。前記学習制御部14は、KDメトリック学習部28にメトリック学習制御信号MLCを送る。

【0162】前記メトリック学習制御信号MLCを受けたKDメトリック学習部28は、まず、1次肯定メトリック記憶部26から1次肯定メトリック信号MY1を、1次否定メトリック記憶部27から1次否定メトリック信号MN1をそれぞれ読み出す。

【0163】次に、KDメトリック学習部28は、教師データカウンタcの値を1にする。

【0164】教師データ記憶部13からc番目の教師データ信号TD[c]を読み出し、教師データTD[c]の教師信号T[c]を調べる。前記教師信号T[c]が-1でない場合(T≠-1)には、教師データTD[c]のキーワード数信号TnofKs[c]とキーワード群信号TKs[c]とを出力する。前記教師データTD[c]のキーワード数信号TnofKs[c]とキーワード群TKs[c]とを受けた学習用ベクトル生成部20は、前述した実施の形態1の情報フィルタリングユニット50のベクトル生成部1と同様の動作を行い、学習用ベクトル信号LVを出力する。KDメトリック学習部28は、前記学習用ベクトル信号LVを受け、前記教師データTD[c]の教師信号T[c]がT=1である場合には、1次肯定メトリック信号MY1を

$$MY1[i][j] = MY1[i][j] + LV[i] \cdot LV[j]$$

(ここで、i, j=1~nofDiC)と修正する。前記教師データTD[c]の教師信号T[c]がT=0である場合には、1次否定メトリック信号MN1を

$$MN1[i][j] = MN1[i][j] + LV[i] \cdot LV[j]$$

(ここで、i, j=1~nofDiC)と修正する。教師データカウンタの値をc=c+1と1だけ増やす。

【0165】以下、KDメトリック学習部28は、同様の動作を、教師データTD[c]の教師信号T[c]がT[c]=-1になるかまたはc=nofTDとなるまで繰り返す。T[c]=-1またはc=nofTDとなると、1次肯定メトリック信号MY1と1次否定メトリック信号MN1の学習を終える。

【0166】次に、回数記憶部25から全肯定回数信号NYと全否定回数信号NNを読み出し、1次肯定メトリ

ック信号MY1と1次否定メトリック信号MN1とを用いて、肯定メトリック信号MYを計算する。

【0167】こうして計算される肯定メトリック信号MY、否定メトリック信号MNは、キーワードコスト信号KDと同様、計算される肯定信号SYと否定信号SNが、ユーザの不要な情報データDが出現する確率

$$NN / (NY + NN)$$

と比較して、文字列W[j]が付いている情報データDがユーザにとって不要である場合の確率

$$PN[j] / (PY[j] + PN[j])$$

が大きく異なる場合に、大きくなるようものであり、ユーザの必要な情報データDが出現する確率

$$NY / (NY + NN)$$

と比較して、文字列W[j]が付いている情報データDがユーザにとって必要である場合の確率

$$PY[j] / (PY[j] + PN[j])$$

が大きく異なる場合に、大きくなるようものであるといった性質を持っていれば、なんでもよい。これを満たす好ましいのは、肯定メトリック信号MYを

【数18】

$$\begin{aligned} MY[i][j] &= NY / (NY + NN) \\ &\cdot \log((MY1[i][j] + e) \cdot (NY + NN)) \\ &/ (NY \cdot (MY1[i][j] + MN1[i][j] + 2e)) \end{aligned}$$

と計算し、否定メトリック信号MNを

【数19】

$$\begin{aligned} MN[i][j] &= NN / (NY + NN) \\ &\cdot \log((MN1[i][j] + e) \cdot (NY + NN)) \\ &/ (NN \cdot (MY1[i][j] + MN1[i][j] + 2e)) \end{aligned}$$

と計算する。ここで、eは0でのわり算、log0を避けるための小さな正の値を持つパラメータである。

【0168】そして、更新された1次肯定メトリック信号MY1を1次肯定メトリック信号記憶部26に、更新された1次否定メトリック信号MN1を1次否定メトリック信号記憶部27に、新たに計算された肯定メトリック信号MYを肯定メトリック記憶部5へ、新たに計算された否定メトリック信号MNを否定メトリック記憶部6へ書き込む。以上で、KDメトリック学習部28は、メトリック学習の処理を終了し、メトリック学習制御信号MLCを学習制御部14に送る。

【0169】学習制御部14は、KDメトリック学習部28からのメトリック学習制御信号MLCを受けて、スイッチ16を学習用ベクトル生成部20とスコア計算部22とが接続するように切り替え、スイッチ17とスイッチ18を学習用ベクトル生成部20と判定面学習部21とが接続するように切り替える。学習制御部14は、判定面学習制御信号PLCを判定面学習部21に送る。

【0170】判定面学習部21の動作は、実施の形態1と全く同じであるので、説明は繰り返さない。

【0171】一度、以上の処理が行われると、辞書記憶部2の符号辞書が空でなくなるので、情報フィルタリングユニット50から出力される必要性信号N、信頼性信号Rは、0でなくなり、ユーザの必要性の高い情報データが、未読データ記憶部10の上位に書き込まれるようになる。

【0172】以後、上記処理を繰り返すことにより、ユーザが必要とする情報が否かを判定するために有効なキーワードが優先的に辞書記憶部2に記憶されるようになり、小規模な辞書であっても、精度の高い情報フィルタリングが可能となる。

【0173】なお、判定パラメータCの計算方法として、ここでは、山登り法を採用したが、実施の形態1と同様、判定面と学習用必要性信号LNと学習用信頼性信号LRとの距離に基づいて構成されるコスト関数を最大にする判定面パラメータCをニュートン法、挟み撃ち法などで求める方法であってもよい。さらに、簡便な方法として、

$$C = \tan \theta_i$$

ここで、

$$\theta_i = 0.5 \cdot \pi (i/90) \quad i = 1, \dots, 90$$

の中から、 $T[c] = 1$ である情報と $T[c] = 0$ である情報をもっともよく分離できるCを選ぶと言う方法も考えることができる。

【0174】また、1次肯定メトリック信号MY1と1次否定メトリック信号MN1の学習を忘却の効果を入れた

$$MY1[i][j] = \alpha \cdot MY1[i][j] + LV[i] \cdot LV[j]$$

$$MN1[i][j] = \alpha \cdot MN1[i][j] + LV[i] \cdot LV[j]$$

を用いてもよい結果が得られる。(ここで、 $\alpha$ は、1より小さい正の数)もしくは、 $MY1[i][j]$ または $MN1[i][j]$ のいずれかが一定値をこえた場合に、

$$MY1[i][j] = MY1[i][j] / 2$$

$$MN1[i][j] = MN1[i][j] / 2$$

として、信号のオーバーフローを防ぐように構成することは、実施上好ましい。これは、適応符号辞書信号FDCK[j]の肯定回数PY[j]と否定回数PN

[j]、および全肯定回数信号NYと全否定回数NNについても同様である。さらに、文献「情報処理学会技術報告、自然言語処理101-8(1994.5.27)」などに記載された文書からキーワード群信号とキーワード数信号を生成するキーワード生成部を付加する構成をとれば、キーワードが与えられていない情報に対しても適用できる情報フィルタ装置を構成することができる。

【0175】タイトルがつけられた情報については、タイトルを構成する単語をもってキーワードとし、キーワード数信号とキーワード群信号を生成してもよい。

【0176】加えて、キーワード信号は、国際特許分類番号など分類記号を含むようにしても、本発明の構成を変更する必要はなく、よい結果をえることができる。

【0177】また、本実施の形態では、未読データURDを一つずつ提示する場合について示したが、表示装置(図示せず)の大きさによっては複数の未読データURDを同時に表示し、ユーザがどの未読データURDについて応答したのかを正しく情報フィルタ装置に伝える構成をとることは容易である。

【0178】また、本発明の実施の形態1では、独立情報だけを表示するように構成したが、ユーザーの要求によって、従属情報も表示できるように構成することは、容易である。

【0179】実施の形態2の情報フィルタ装置によれば、情報相互の類似度を計算して類似情報をひとまとめにするとともに、代表情報及び独立情報を一覧表示するようにしたので、ユーザーに必要性の順に情報を提示する場合に、同じ内容の情報が何度も繰り返し提示されるといった不具合を防止できる。

【0180】実施の形態2の情報フィルタ装置によれば、キーワードの同時出現に注目したメトリックを導入することにより、キーワードという記号情報を距離の定義された空間に射影したことにある。これによって、キーワード群の遠近を距離というアナログ尺度で評価することができるようになる。これを利用することにより、従来の技術では必要か不要かの二者択一的な判定しかできなかった必要性の評価が、ユーザーの必要性の順番に並べるといったことが可能になる。

【0181】本実施の形態による情報フィルタによれば、ユーザーからの教師信号に基づいた学習によって、ユーザーの必要とする情報に対しては、必要性信号が大きな値を取るようになり、その結果、表示装置等には、ユーザーにとって必要性が高い情報が優先的に、かつ類似の情報は代表的なものが一つ表示されるようになる。

【0182】

【発明の効果】以上のように、本発明は、情報記録媒体又は情報通信網から供給された情報が記憶される記憶部と、該記憶部の情報を類似する情報でまとめて代表情報と従属情報を決める類似判定手段と、前記記憶部の情報を提示する場合に類似する情報については代表情報を提示する提示手段とを具備することで、同じような情報が複数含まれる場合には、代表的な情報と関連情報の件数表示することにより、できる限り多くの種類の情報を一度に表示でき一覧性の高い情報フィルタリングを実現できる。

【図面の簡単な説明】

【図1】本発明の実施の形態1の情報フィルタ装置のブ



## ロック結線図

【図2】本発明の実施の形態1の情報フィルタ装置の概略を示すブロック結線図

【図3】本発明の実施の形態1の情報フィルタ装置のベクトル生成部の動作を説明するフローチャート

【図4】本発明の実施の形態1の情報フィルタ装置の未読データ書き込み制御部の動作を説明するフローチャート

【図5】本発明の実施の形態1の情報フィルタ装置の未読データ出力制御部の動作を説明するフローチャート

【図6】本発明の実施の形態1の情報フィルタ装置の学習制御部の動作を説明するフローチャート

【図7】本発明の実施の形態1の情報フィルタ装置のメトリック学習部の動作を説明するフローチャート

【図8】本発明の実施の形態1の情報フィルタ装置の判定面学習部の動作を説明するフローチャート

【図9】本発明の実施の形態1の情報フィルタ装置の判定面学習部の動作を説明するための図

【図10】本発明の実施の形態1の情報フィルタ装置の判定面学習部の動作を説明するための図

【図11】本発明の実施の形態2の情報フィルタ装置のブロック結線図

【図12】本発明の実施の形態2の情報フィルタ装置の辞書学習部の動作を説明するフローチャート

【図13】本発明の実施の形態1の未読データ圧縮部の概略を示すブロック結線図

【図14】本発明の実施の形態1の未読データ圧縮部の動作の前半を説明するためのフローチャート

【図15】本発明の実施の形態1の未読データ圧縮部の動作の後半を説明するためのフローチャート

【図16】本発明の実施の形態1の類似情報リスト信号ADの内容を示す概念図

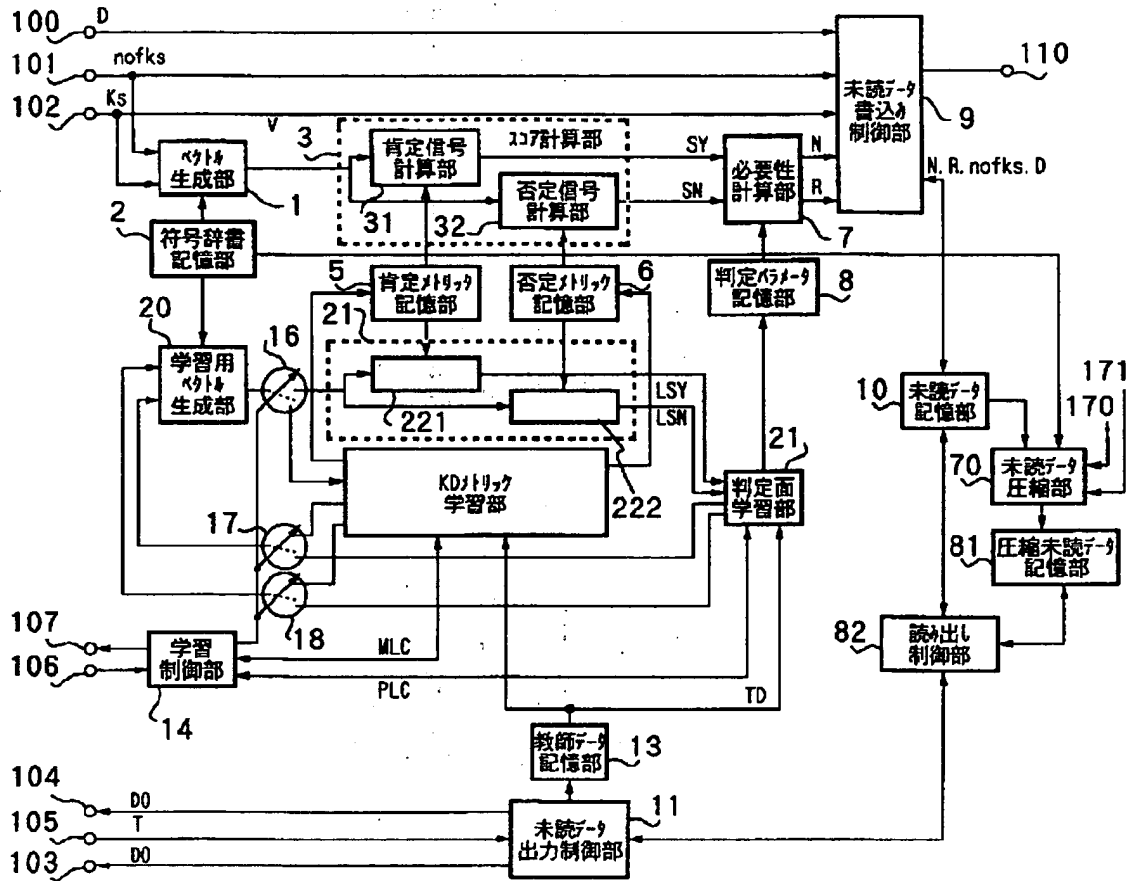
【図17】(a), (b)は本発明の実施の形態1の類似度の特徴を説明するための2つの情報に付されたキーワードの包含関係を示す概念図

【符号の説明】

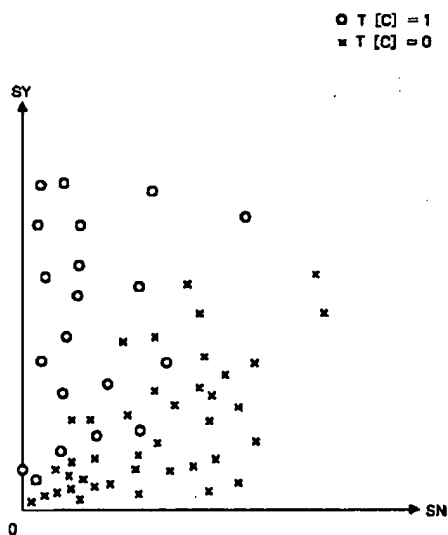
- 1 ベクトル生成部
- 2 符号辞書記憶部
- 3 スコア計算部
- 5 肯定メトリック記憶部
- 6 否定メトリック記憶部
- 7 必要性計算部
- 8 判定パラメータ記憶部
- 9 未読データ書き込み制御部
- 10 未読データ記憶部

- 11 未読データ出力制御部
- 12 教師データ制御部
- 13 教師データ記憶部
- 14 学習制御部
- 16 スイッチ
- 17 スイッチ
- 18 スイッチ
- 19 メトリック学習部
- 20 学習用ベクトル生成部
- 21 判定面学習部
- 22 スコア計算部
- 23 辞書学習部
- 24 適応符号辞書記憶部
- 25 回数記憶部
- 26 1次肯定メトリック記憶部
- 27 1次否定メトリック記憶部
- 28 KDメトリック学習部
- 31 肯定信号計算部
- 32 否定信号計算部
- 50 情報フィルタリングユニット
- 51 インタフェースユニット
- 52 学習ユニット
- 53 代表情報一覧ユニット
- 70 未読データ圧縮部
- 71 類似度計算制御部
- 72 未読データ読み出し部
- 73 類似度ベクトル生成部
- 74 スイッチ
- 75 参照ベクトル記憶部
- 76 類似度計算部
- 77 閾値記憶部
- 78 比較部
- 79 圧縮未読データ書き込み制御部
- 80 フラグ記憶部
- 81 圧縮未読データ記憶部
- 82 圧縮未読データ読み出し制御部
- 101 キーワード数信号入力端子
- 102 キーワード信号入力端子
- 103 データ読み出し開始信号入力端子
- 104 データ表示端子
- 105 教師信号入力端子
- 106 学習開始信号入力端子
- 107 学習制御部指示信号出力端子
- 110 未読データ部指示端子

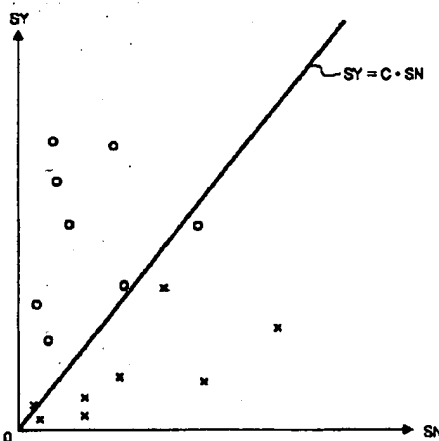
【図1】



【図9】



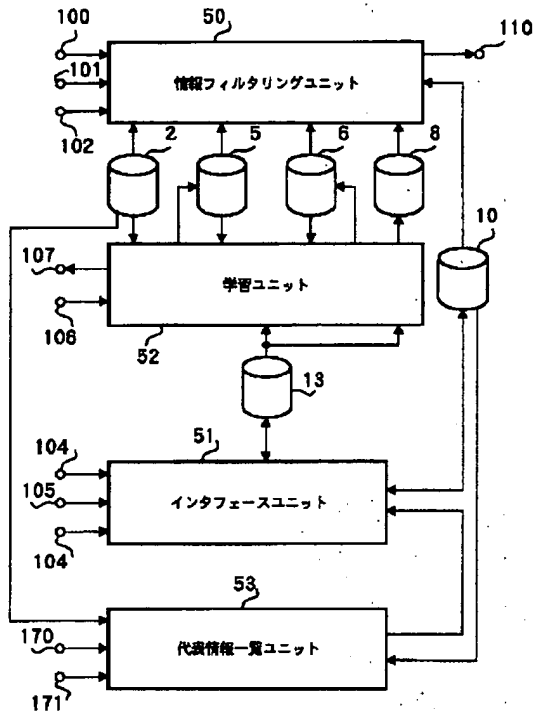
【図10】



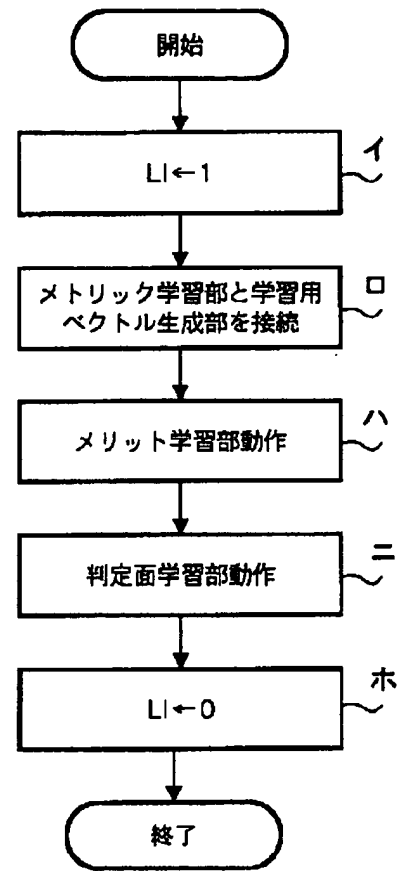
【図16】

1	8	5	-	-	-	-
2	-	-	-	-	-	-
4	6	7	-	-	-	-
8	10	-	-	-	-	-
9	-	-	-	-	-	-
-	-	-	-	-	-	-

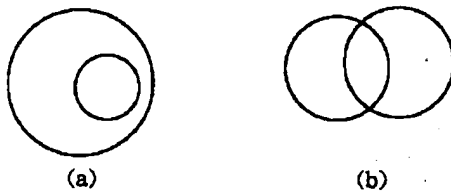
【図2】



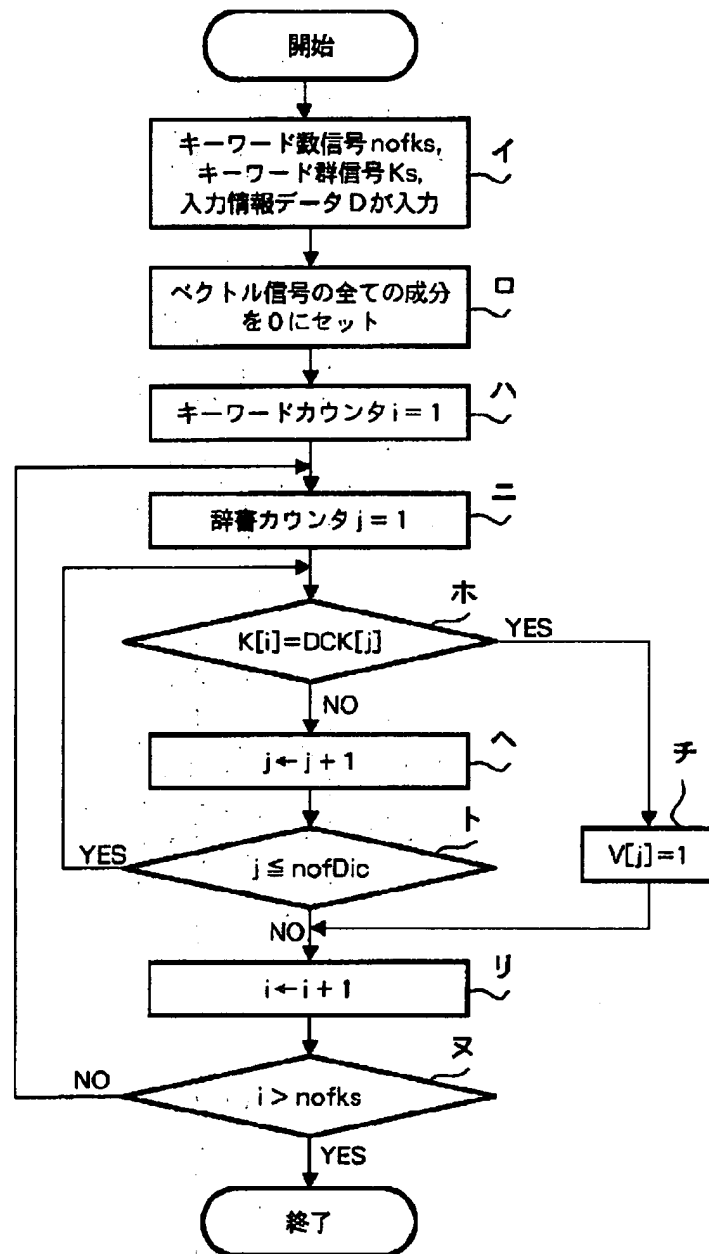
【図6】



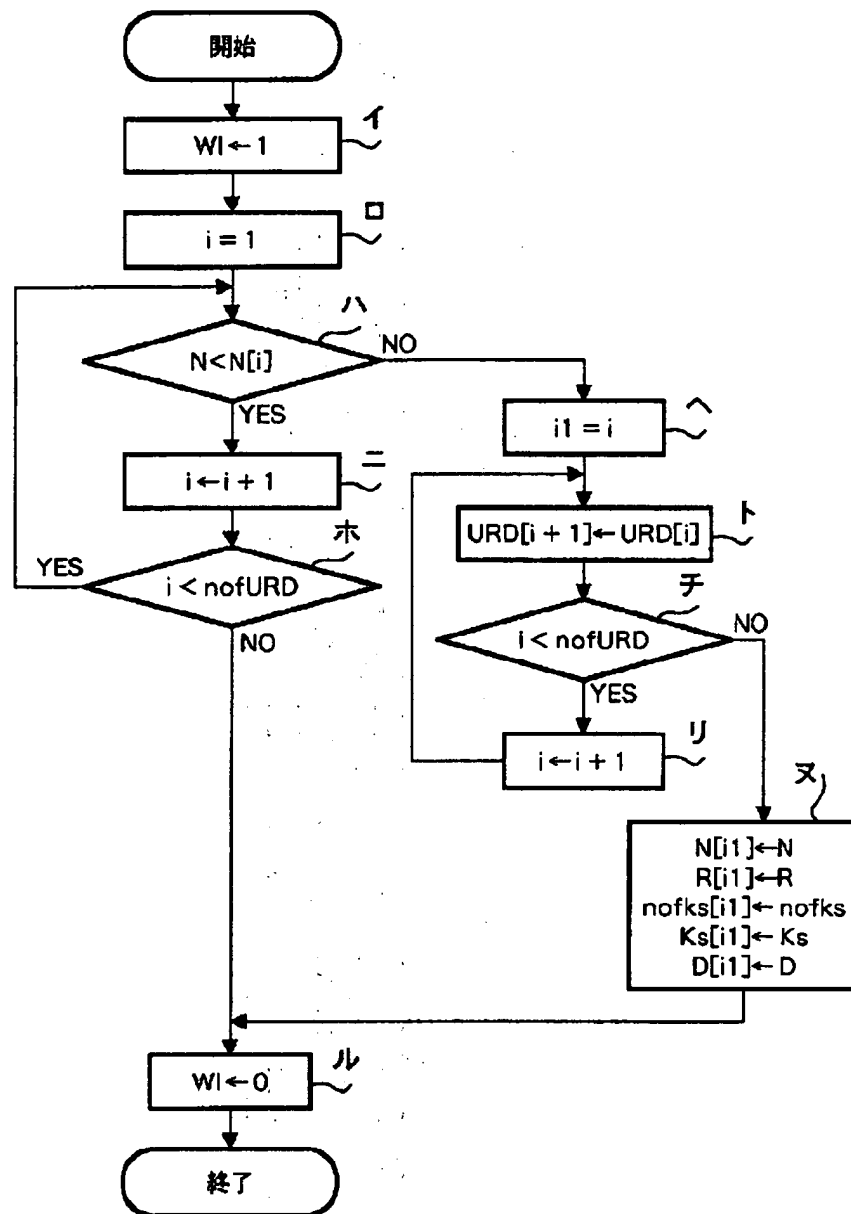
【図17】



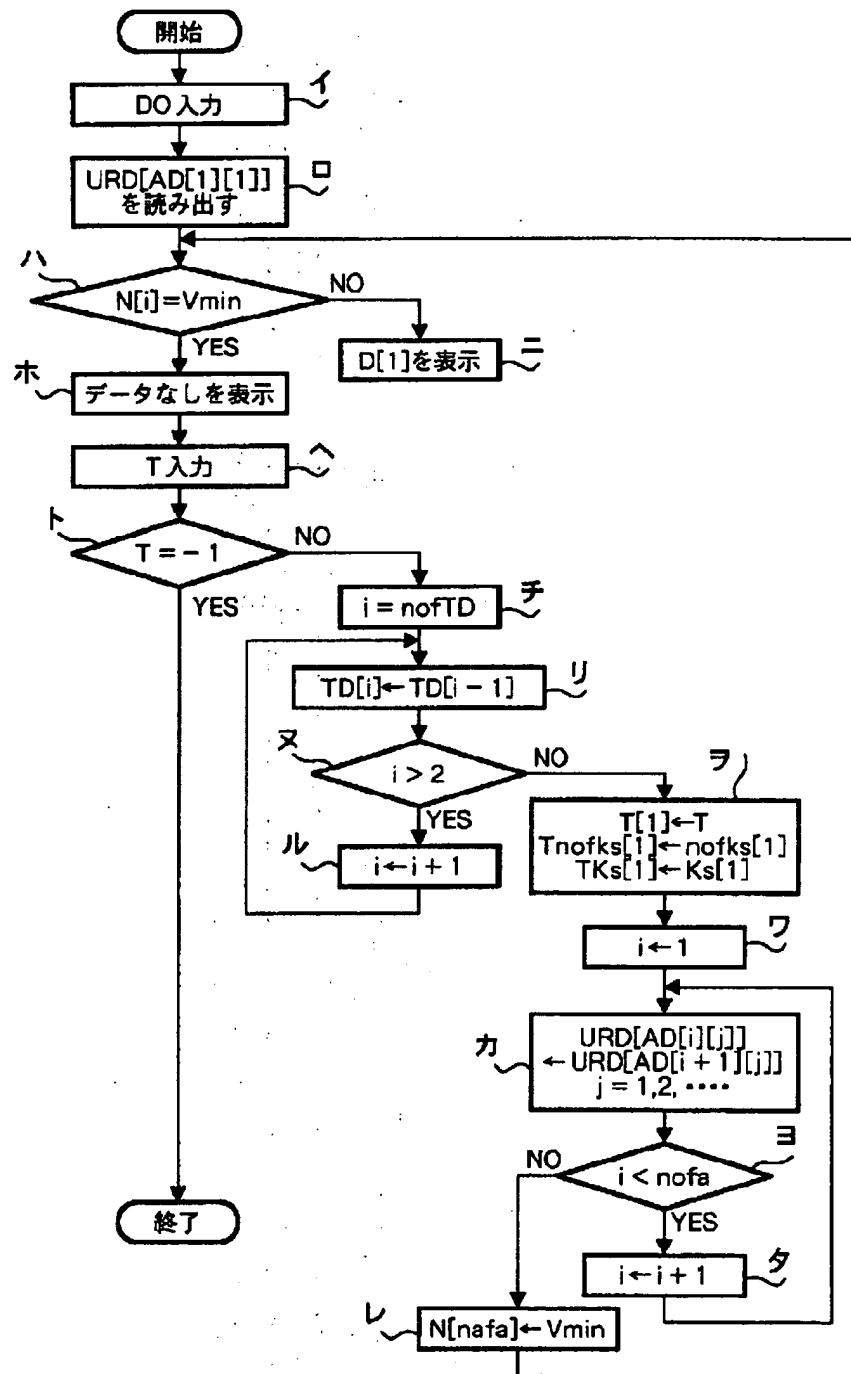
【図3】



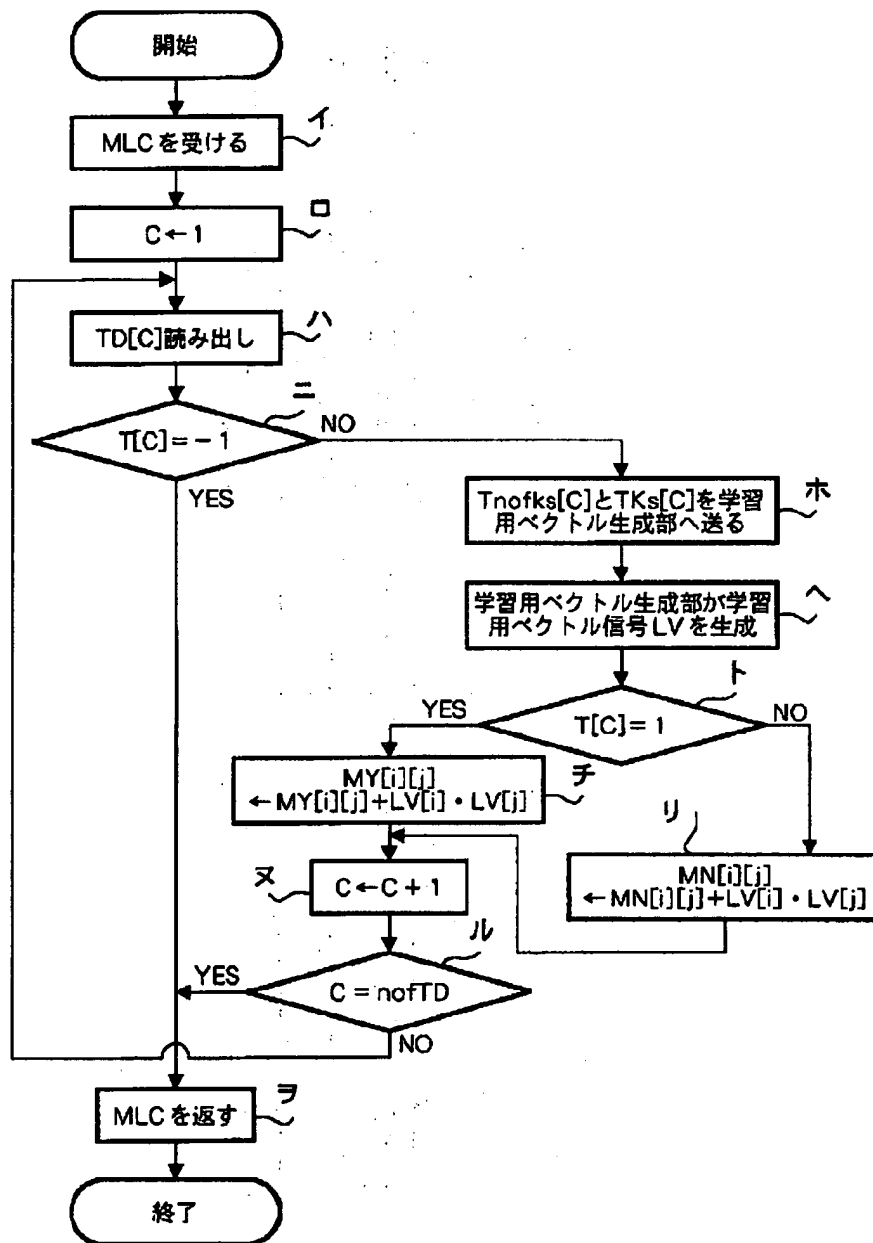
【図4】



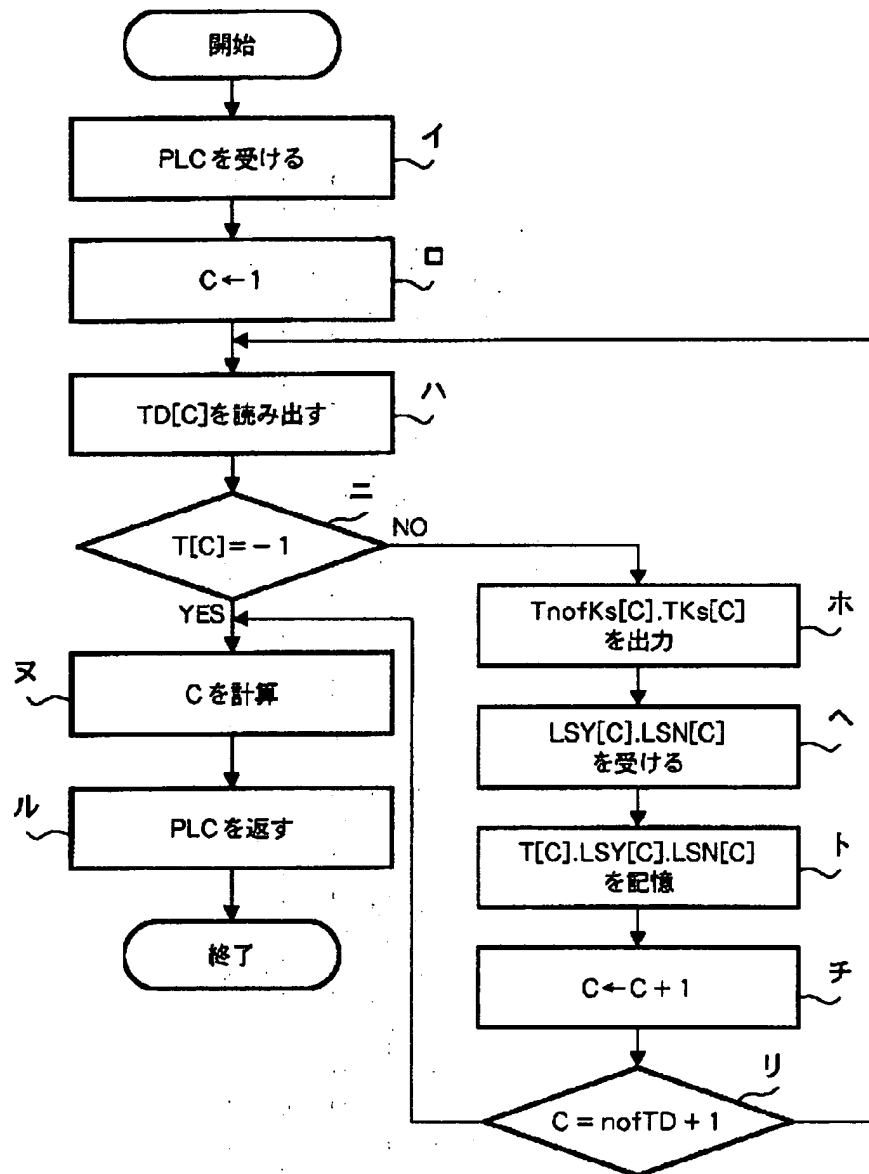
【図5】



【図7】

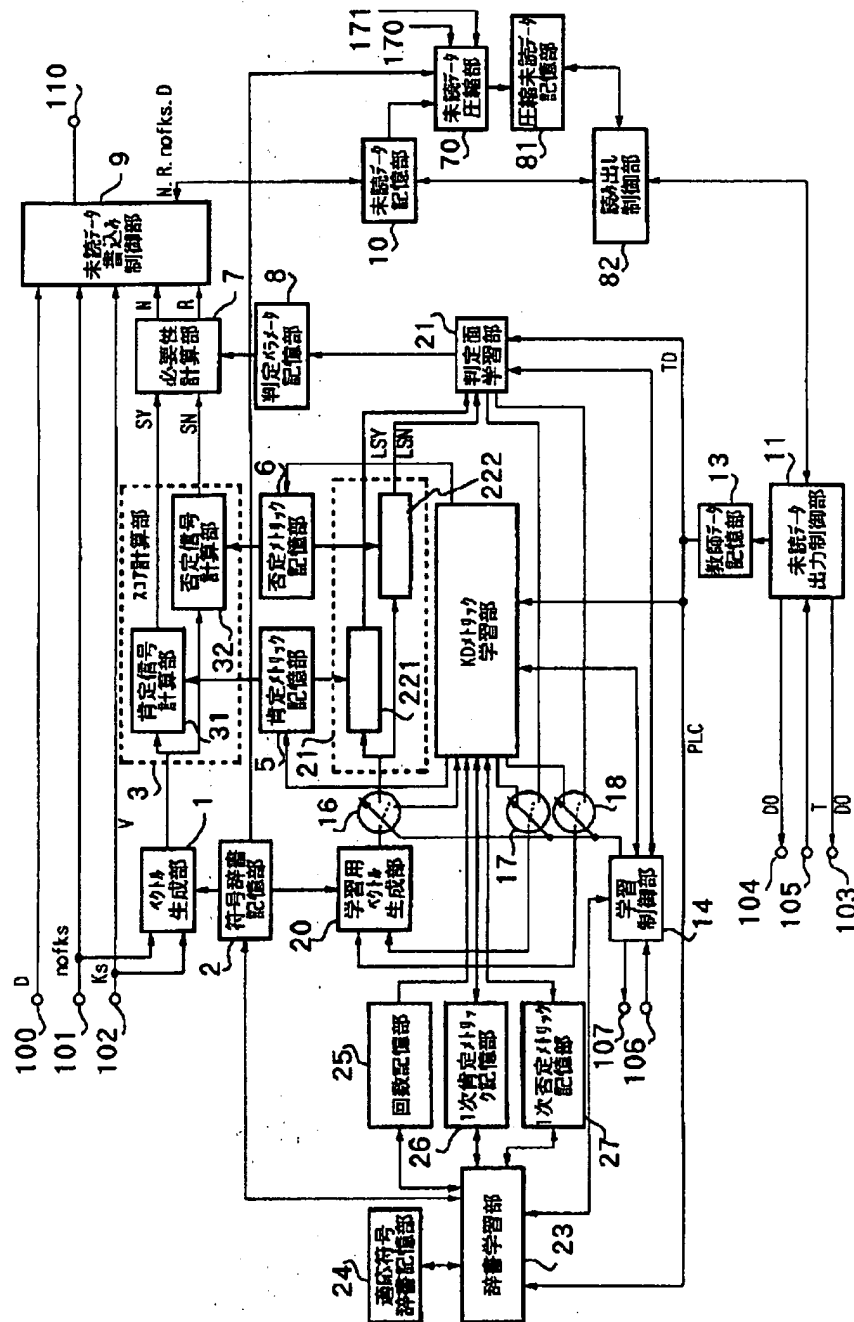


【図8】

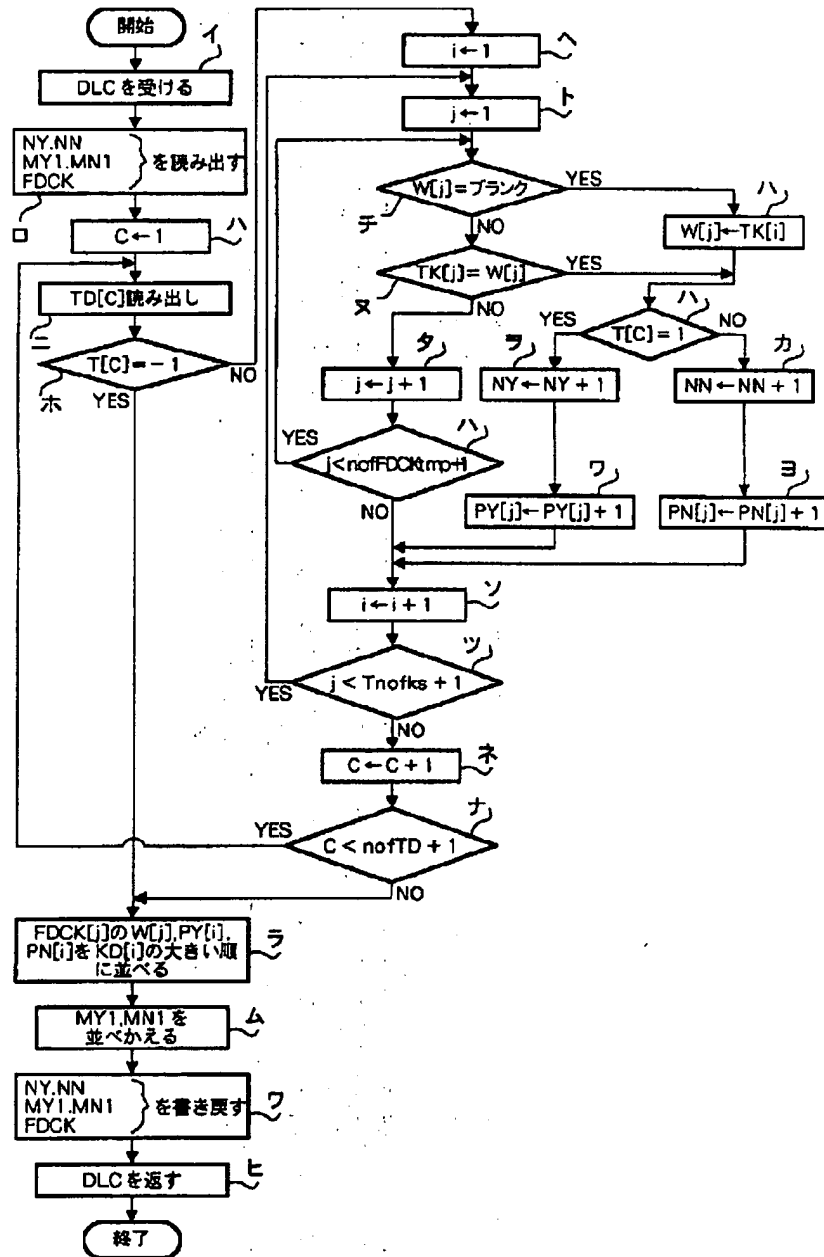




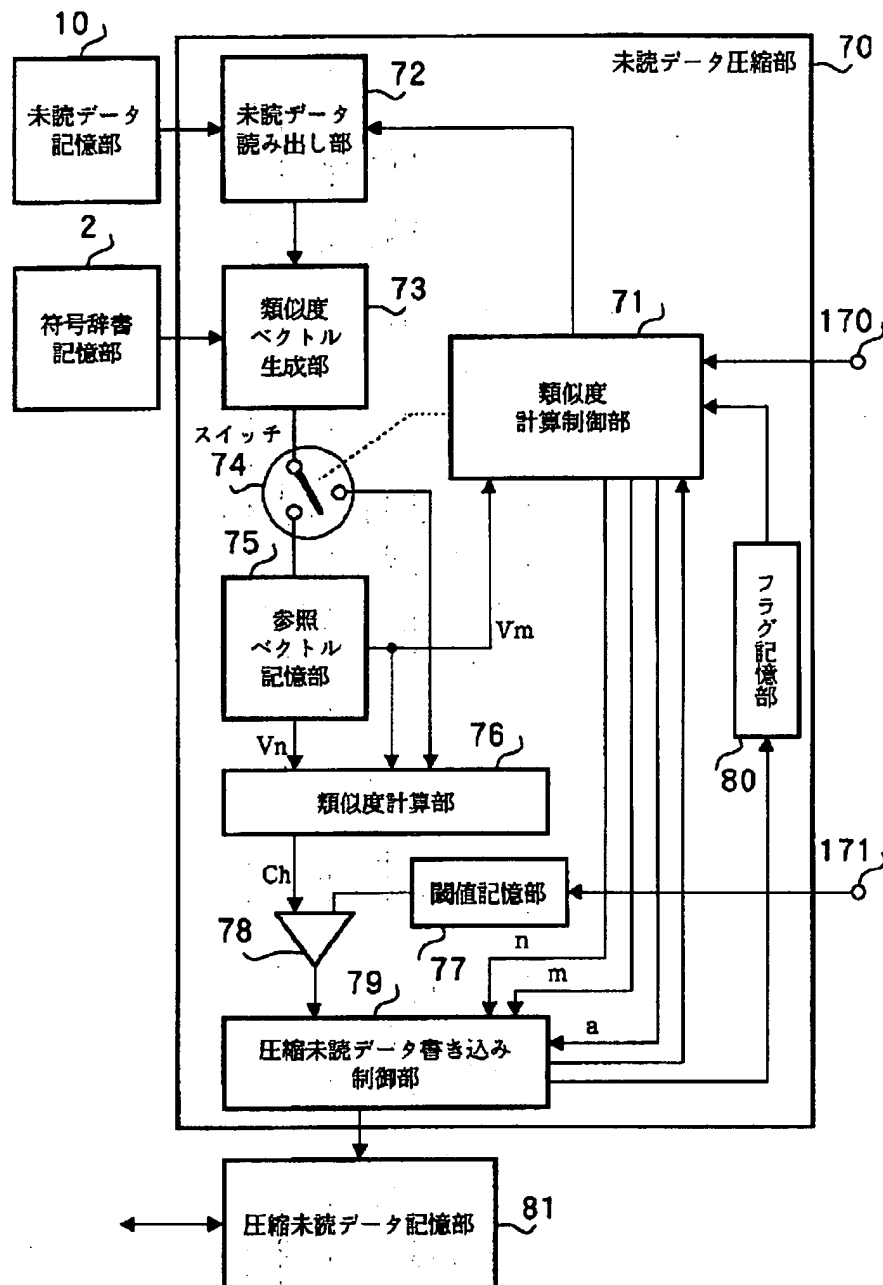
【図 1 1】



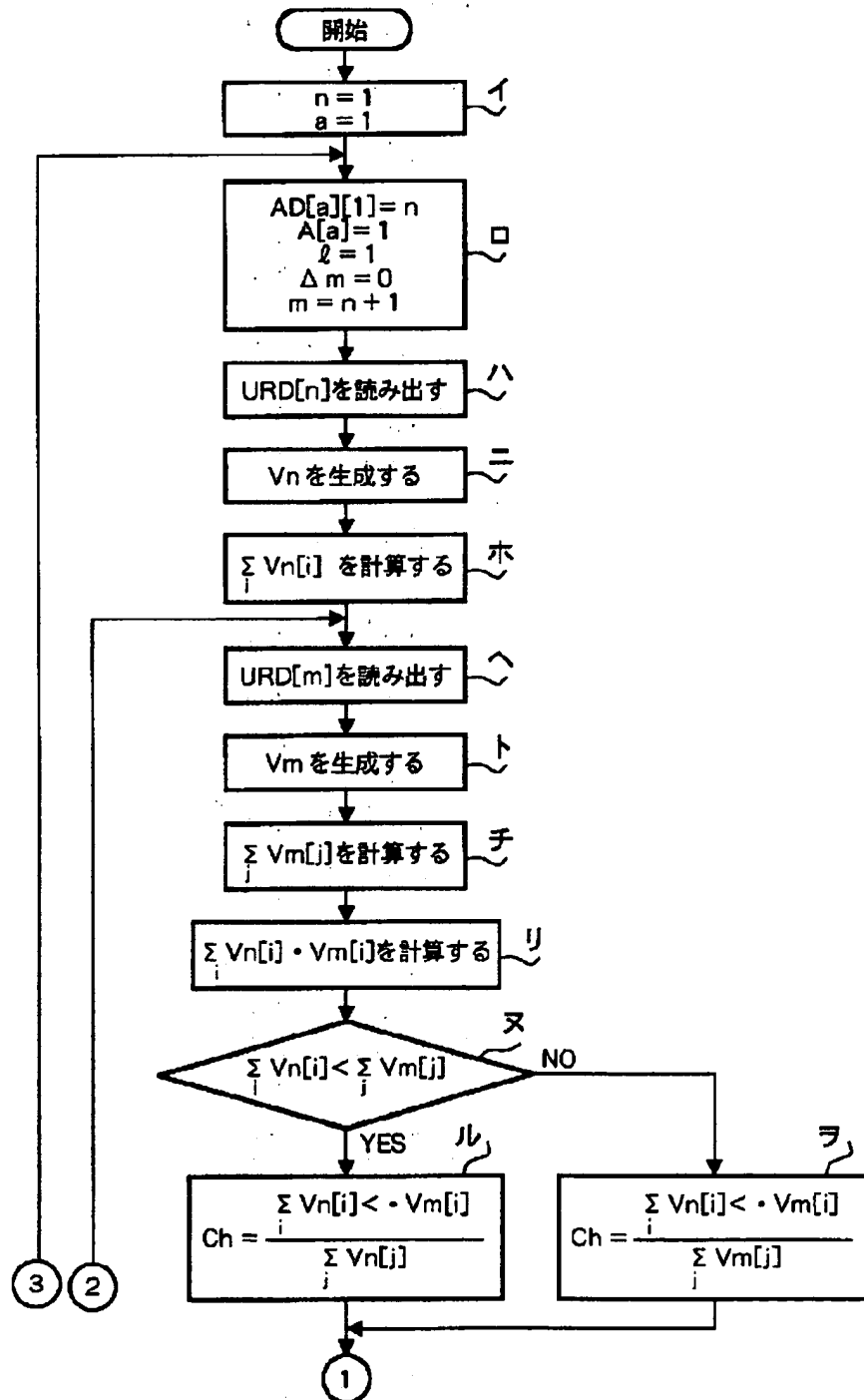
【図12】



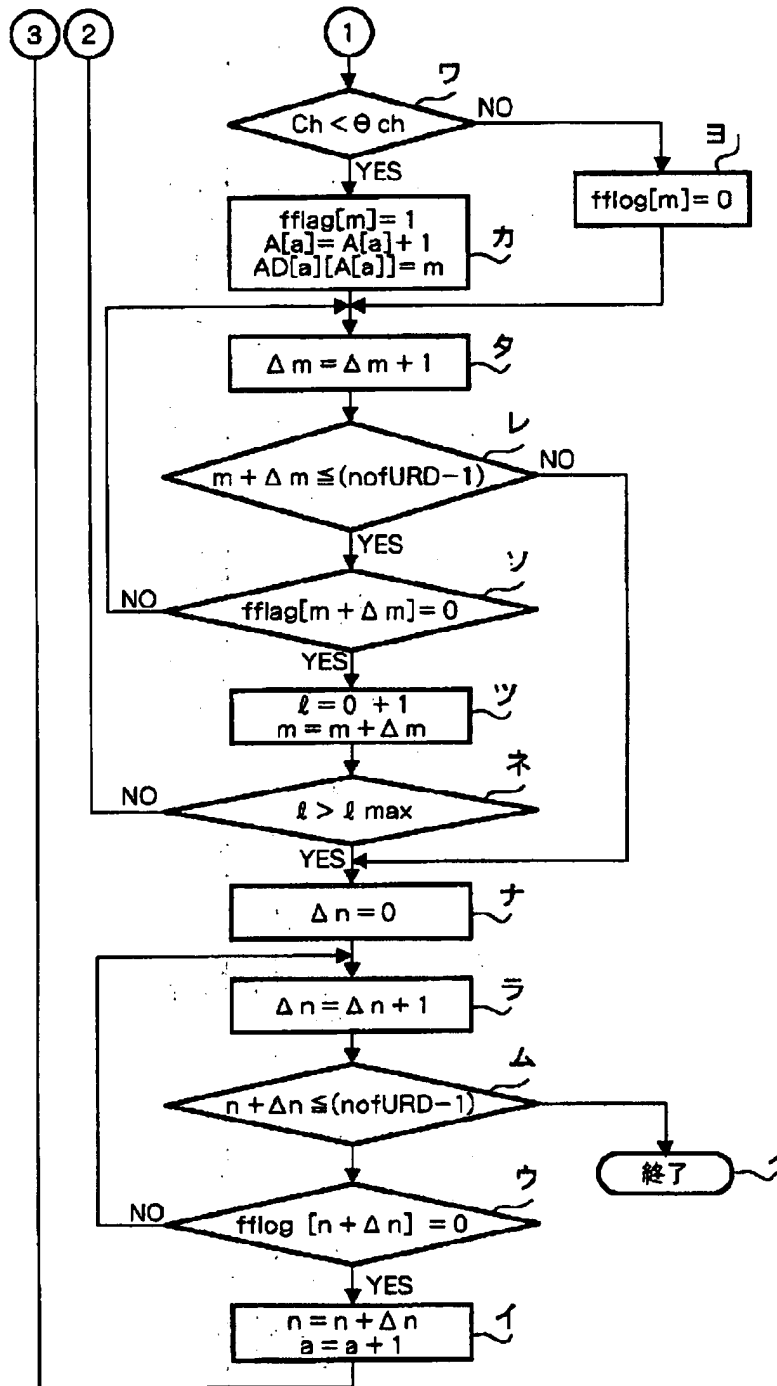
【図13】



【図14】



【図15】



## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-275160

(43)Date of publication of application : 13.10.1998

(51)Int.Cl.

G06F 17/30  
// G06F 15/18

(21)Application number : 09-096418

(71)Applicant : MATSUSHITA GIKEN KK

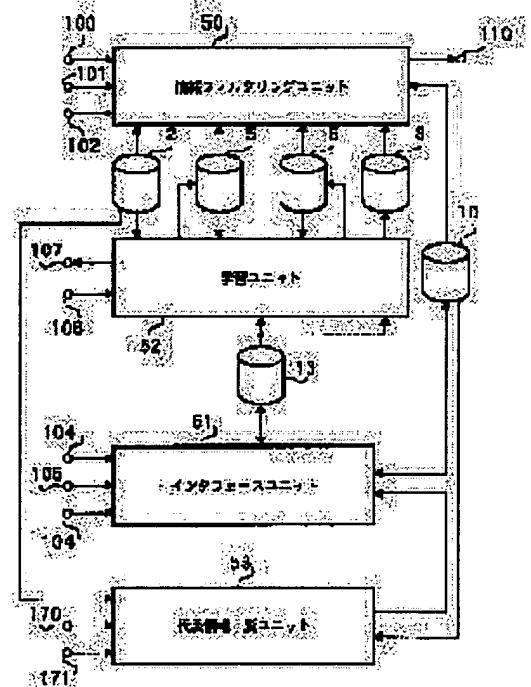
(22)Date of filing : 31.03.1997

(72)Inventor : KANEMICHI TOSHIKI  
YOSHIDA HIDEYUKI  
WATANABE TAISUKE(54) INFORMATION FILTER DEVICE, INFORMATION FILTERING METHOD, AND RECORDING MEDIUM  
RECORDED WITH INFORMATION FILTERING PROGRAM

## (57)Abstract:

**PROBLEM TO BE SOLVED:** To display information as much as possible at a time by providing a means for determining representative information and subordinate information by putting together similar pieces information in a storage part and a means for presenting the representative information with respect to similar information when information in the storage means is displayed.

**SOLUTION:** A representative information list unit 53 calculates mutual similarity of information stored in the storage part 10 by an unread data compression part when writing to the unread data storage part 10 is completed, and writes a result showing which data can be put together in a compressed unread data storage part. An interface unit 51 shows unread information including new information and representative information when there is similar information to a user in the decreasing order of necessity signals. Then a learning unit 52 rewrites history contents of storage parts 2, 5, 6, and 8 by using the shown information and its tutor signal. Consequently, information that the user requires can be shown preferentially while listed with precision.



## LEGAL STATUS

[Date of request for examination] 24.07.1998

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the  
examiner's decision of rejection or application converted  
registration]

[Date of final disposal for application]

[Patent number] 3203203

[Date of registration] 22.06.2001

[Number of appeal against examiner's decision of  
rejection][Date of requesting appeal against examiner's decision of  
rejection]

**\* NOTICES \***

**Japan Patent Office is not responsible for any damages caused by the use of this translation.**

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

---

**CLAIMS**


---

[Claim(s)]

[Claim 1] It is information filter equipment possessing the storage section the information supplied from the information record medium or the information communication network is remembered to be, a similar judging means summarize the information on this storage section for similar information, and decide representation information and subordinate information, and a presentation means show representation information about the information which is similar when showing the information on the aforementioned storage section.

[Claim 2] The storage section is information filter equipment according to claim 1 characterized by putting information in order according to the height of the need of having calculated two or more keywords assigned by information and the information to which these keywords were assigned in the past based on the need / history made unnecessary.

[Claim 3] A similar judging means is information filter equipment according to claim 1 or 2 characterized by searching information similar to representation information based on the degree of coincidence of the keyword attached to each information memorized by the storage section.

[Claim 4] A presentation means is information filter equipment according to claim 1 to 3 characterized by showing representation information and independent information in the high sequence of need.

[Claim 5] A presentation means is information filter equipment according to claim 1 to 4 characterized by displaying the number or existence of subordinate information similar to the representation information concerned together in case representation information is shown.

[Claim 6] A similar judging means is information filter equipment according to claim 1 to 5 characterized by to search the subordinate information which makes the independent information representation information and is similar in the direction of a low of need from the representation information concerned if there is independent information which checks the information on the storage section in an order from the high information on need, and is not yet representation information or subordinate information.

[Claim 7] A similar judging means is information filter equipment according to claim 6 characterized by directing or setting up beforehand the reference range which searches similar information and similar independent information, and shifting the reference range concerned in the direction of a low of need one by one.

[Claim 8] Information filter equipment according to claim 1 to 7 characterized by providing the following A means to change into a vector signal two or more keyword signals assigned by information The score calculation section which calculates an affirmation score and a negative score using the affirmation metric signal which consists of required information, the negative metric signal which consists of unnecessary information, and the aforementioned vector signal The need [ of calculating a need signal and a reliability signal from the judgment parameter signal which divides into required information and unnecessary information the distribution of the point of a 2-dimensional flat surface which consists of the aforementioned affirmation score signal and the aforementioned negative score signal, the aforementioned affirmation score signal, and the aforementioned negative score signal ] calculation section The means which arranges information in order of the size of the aforementioned need signal, and is written in the aforementioned storage section

[Claim 9] It is the information filtering method characterized by showing representation information about the information which is similar when summarizing the information supplied from the information record medium or the information communication network for similar information, deciding representation information and subordinate information and showing the aforementioned information.

[Claim 10] It is the information filtering method according to claim 9 characterized by to take out information from the storage section with which information was compared in the turn according to the height of need, to search subordinate information similar to representation information based on the degree of coincidence of the keyword attached to information, and to show information in the high sequence of need by making representation information applicable to

presentation about similar information.

[Claim 11] It is the record medium which recorded the information filtering program information makes a program show in the high sequence of need by making representation information applicable to presentation about the information which is the record medium which recorded the information filtering program which filters information by computer, and an information filtering program makes detect the subordinate information which is similar to representation information from the information supplied to the computer from the information record medium or the information communication network, and is similar.

---

[Translation done.]



**\* NOTICES \***

**Japan Patent Office is not responsible for any damages caused by the use of this translation.**

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

**DETAILED DESCRIPTION****[Detailed Description of the Invention]**

**[0001]**

[The technical field to which invention belongs] this invention relates to the record medium which recorded the information filter equipment, the information filtering method, and information filtering program make required information easy to take out from the storage through an electron or light, or an information communication network.

**[0002]**

[Description of the Prior Art] As technology for coping with large-scale-izing of the information communication network accompanying development of an information communication base in recent years, and remarkable increase of traffic, realization of information filter equipment is desired strongly. For this background, the amount of information which an individual can access may exceed to the amount of information which an individual can process today. For this reason, it often happens that information required for a user in a lot of information is buried.

[0003] As conventional technology relevant to information filter equipment, the keyword logical expression used for patent reference etc. can be held. That is, the patent information which attains to 1 million numbers from hundreds of thousands is filtered by keyword logical expression.

[0004] However, in the conventional reference using keyword logical expression, since a user needs to set up the logical expression about a keyword with a sufficient precision, if the peculiarity (for example, the keyword of the data concerned is determined as the basis of what conditions) of a data constellation and the structure (for example, \*\*\*\*\* [ that a keyword is a system with a thesaurus system ] etc.) of a system where the user is filed can fully be known, good reference cannot be performed. For this reason, the beginner had the technical problem that information filtering with a high precision could not be performed.

[0005] Moreover, although the result which carried out information filtering only has evaluation of suiting the logical expression about a keyword and it has agreed by the keyword by chance, it is not easy for the content to have a different case from that for which it is asking, or to take out the information that the degree of required is high for a user from many reference results sequentially from the result.

[0006] Then, information filtering with a high precision is possible also to a beginner, and for a user, this invention persons develop the information filter equipment which can take out the high information on need easily, and are patent application ending as Japanese Patent Application No. No. 230012 [ eight to ], and Japanese Patent Application No. No. 046384 [ nine to ].

[0007] It calculates informational need and informational reliability by this information filter equipment changing into a vector two or more keywords assigned by information, and calculating a score using the matrix expressing whether the vector and the user needed what information and it was presupposed that it is unnecessary. Information is rearranged into descending of need, and information will be shown to descending of need if there is a demand from a user. Furthermore, the shown information inputs evaluation of the need or the user whether to be unnecessary, and corrects the matrix to which have and it is in score calculation from evaluation of a user and two or more aforementioned keywords.

[0008] By this composition, two or more keywords can be changed into the vectorial representation which can define distance from the sign which cannot perform the definition of distance using metric one reflecting a user's degree of required, and can quantify a user's degree of required, and a user can acquire information now sequentially from the high information on need.

**[0009]**

[Problem(s) to be Solved by the Invention] However, in the above-mentioned information filter equipment, when information was collected from two or more information sources (for example, two or more newspaper publishing companies), the information on the same content appeared repeatedly and the technical problem that the list nature of

sufficient information was not obtained occurred.

[0010] For example, a high score will be attached to the major accident (information) interesting also for a user when a major accident which is reported in respect of many newspapers occurs. Consequently, if information is shown to descending of need, the content of information presentation is occupied with the report of the major accident (information), and by the time it acquires the information on other, the report of the same content will be shown repeatedly.

[0011] this invention aims at being able to display as many kinds as possible of information at once, and offering the high information filter equipment and the information filtering method of list nature by [ typical / of information and related information ] indicating by the number of cases, when it is made in view of the above actual condition and two or more same information is included.

[0012]

[Means for Solving the Problem] In order to attain the above-mentioned technical problem, this invention provided the following meanses.

[0013] Invention according to claim 1 takes the composition possessing the storage section the information supplied from the information record medium or the information communication network is remembered to be, a similar judging means summarize the information on this storage section for similar information, and decide representation information and subordinate information, and a presentation means show representation information about the information which is similar when showing the information on the aforementioned storage section.

[0014] Invention according to claim 9 summarizes the information supplied from the information record medium or the information communication network for similar information, determines representation information and subordinate information, and takes the composition which presents representation information about the information which is similar when showing the aforementioned information.

[0015] Since representation information is shown about the information which is similar when showing the information on the storage section according to this invention, many kinds of information can be displayed at once, and the improvement of list nature can be aimed at.

[0016] Invention according to claim 2 takes the composition which puts in order two or more keywords assigned by information in the information on the storage section, and the information to which these keywords were assigned in the past in the sequence according to the height of the need of having been calculated based on the need / history made unnecessary.

[0017] Since the information on the storage section is put in order in the sequence according to the height of need, while being able to show information in the turn according to liking of a user according to this invention, only the representation information represented about similar information can be displayed, and the improvement of list nature can be aimed at.

[0018] Invention according to claim 3 takes the composition which searches the information for which a similar judging means is similar to representation information based on the degree of coincidence of the keyword attached to each information memorized by the storage section.

[0019] Since representation information is shown about the information which is similar when according to this invention information similar to representation information based on the degree of coincidence of a keyword can be judged and it shows the information on the storage section, many kinds of information can be displayed at once, and the improvement of list nature can be aimed at.

[0020] Invention according to claim 4 takes the composition which presents representation information in the high sequence of need.

[0021] Invention according to claim 10 takes out information from the storage section with which information was compared in the turn according to the height of need, and takes the composition which searches subordinate information similar to representation information based on the degree of coincidence of the keyword attached to information, and presents information in the high sequence of need by making representation information applicable to presentation about similar information.

[0022] According to this invention, since representation information is shown in the high sequence of need, information can be shown in the turn according to liking of a user, only the representation information represented about similar information can be displayed, and the improvement of list nature can be aimed at.

[0023] In case invention according to claim 5 presents representation information, it takes the composition which displays the number or existence of subordinate information similar to the representation information concerned together.

[0024] More related information can be known securing list nature according to this invention, since the number or existence of subordinate information is displayed with representation information.

[0025] If invention according to claim 6 has the independent information which checks the information on the storage section in an order from the high information on need, and is not yet representation information or subordinate information, it will take the composition which searches the subordinate information which makes the independent information representation information and is similar in the low direction of need from the representation information concerned.

[0026] According to this invention, since information is located in a line in the high sequence of importance, the increase in efficiency of the reference range at the time of searching similar information can be attained using similar information not existing above representation information.

[0027] The composition which the reference range which searches the information and the independent information that invention according to claim 7 is similar is directed or set up beforehand, and makes shift the reference range concerned in the low direction of need one by one is taken.

[0028] If according to this invention the reference range is defined when information is located in a line in the high sequence of importance, the increase in efficiency of reference can be attained using similar information having the property which importance approximates and existing in a comparatively near position.

[0029] A means to change into a vector signal two or more keyword signals with which invention according to claim 8 was assigned by information, The score calculation section which calculates an affirmation score and a negative score using the affirmation metric signal which consists of required information, the negative metric signal which consists of unnecessary information, and the aforementioned vector signal, The judgment parameter signal which divides into required information and unnecessary information the distribution of the point of a 2-dimensional flat surface which consists of the aforementioned affirmation score signal and the aforementioned negative score signal, The composition possessing the need [ of calculating a need signal and a reliability signal from the aforementioned affirmation score signal and the aforementioned negative score signal ] calculation section, and the means which arranges information in order of the size of the aforementioned need signal, and is written in the aforementioned storage section is taken.

[0030] According to this invention, two or more keywords can be changed into the vectorial representation which can define distance using metric one reflecting a user's degree of required from the sign which cannot perform the definition of distance, a user's degree of required can be quantified, and a user can acquire information now sequentially from the high information on need.

[0031] Invention according to claim 11 is the record medium which recorded the information filtering program which filters information by the computer, and an information filtering program takes the composition which records the information filtering program make the subordinate information which is similar to representation information from the information supplied to the computer from the information record medium or the information communication network detect, and information makes a program show in the high sequence of need by making representation information applicable to presentation about similar information.

[0032] Analogous information is summarized to typical information and comes to be able to perform high information filtering of list nature according to this invention.

[0033]

[Embodiments of the Invention] Hereafter, the form of operation of this invention is explained in detail with reference to a drawing.

[0034] (Form 1 of operation) Drawing 1 is the block diagram showing the composition of the information filter equipment concerning the form 1 of operation, and drawing 2 is the block diagram which packed information filter equipment into the functional unit.

[0035] This information filter equipment The user in the past what "information" It is based on the history whether to have needed. "informational" filtering The interface units 51, such as a display which shows a user the information filtering unit 50 to perform and the filtered information, the study unit 52 which performs study about the history what "information" the user needed, the filtered unread data It is divided into the representation information list unit 53 made the "informational" list which collects into the unread data with which the contents were well alike from (the information which the user is not yet reading), and represents each group.

[0036] two or more keywords (correctly) by which the information filtering unit 50 was attached to each "information" The portion which calculates the affirmation signal in which a user expresses the score of the kind which has used the affirmation metric signal and negative metric signal which expressed the need / history of having presupposed that it is unnecessary for what "information" as the portion which changes the character string containing a classification code into a vector, and a negative signal, It consists of a portion which calculates the need [ of often reflecting the need for "informational" from this affirmation signal and a negative signal ] signal, and a portion which rearranges information into descending of this need signal.

[0037] In drawing 1 , the sign dictionary signal for the vector generation section 1 changing into a vector two or more

character strings, such as a keyword attached to "information", and the sign dictionary storage section 2 changing two or more character strings, such as a keyword, into a vector is memorized. the correspondence table which changes into Number C the character strings W, such as a keyword which attaches to "information" the sign dictionary signal memorized by this sign dictionary storage section 2, -- a nofDCK piece -- the code book [several 1] which it has

$$\begin{aligned} & \vdots \\ & \text{DCK}[\text{nofDCK}] = (\text{W}[\text{nofDCK}], \text{C}[\text{nofDCK}]) \end{aligned}$$

It comes out. The vector generation section 1 receives keyword group signal  $K_s = (K[1], \dots, K[\text{nofKs}])$  which consists of a number signal nofKs of keywords, and a keyword signal of a nofKs individual, and changes it into the vector signal V using the keyword group signal  $K_s$  and the aforementioned sign dictionary signal DCK. Moreover, the score calculation section 3 changes the "information" which the user was shown into the length of two vector signals V and the affirmation signal SY which were changed in the vector generation section 1, and the negative signal SN using the affirmation metric signal MY calculated from the need / result it was estimated that was unnecessary, and the negative metric signal MN. The affirmation metric storage section 5 memorizes the aforementioned affirmation metric signal MY which is a matrix (nofDCKxnofDCK), the negative metric storage section 6 memorizes the aforementioned negative metric signal MN which is a matrix (nofDCKxnofDCK), and the judgment parameter storage section 8 is a portion which memorizes the judgment parameter signal C. The need calculation section 7 receives the aforementioned affirmation signal SY and the aforementioned negative signal SN, reads the judgment parameter signal C from the aforementioned distinction parameter storage section 8, and calculates the need signal N and the reliability signal R. Furthermore, it writes in the unread data-storage section 10 in which the unread data write-in control section 9 mentions later the information data D, and the number signal nofKs of keywords and the keyword group signal  $K_s$  which is the "informational" text, the need signal N, and the reliability signal R according to a predetermined procedure. The unread data-storage section 10 is unread data [several 2] of a maximum nofURD individual which consist of the information data D, and the aforementioned number signal nofKs of keywords and the aforementioned keyword group signal  $K_s$  which is the text of the above "information", the aforementioned need signal N, and the aforementioned reliability signal R.

$$\begin{aligned} & \text{URD}[1] = (\text{N}[1], \text{R}[1], \text{nofKs}[1], \text{Ks}[1], \text{D}[1]) \\ & \vdots \\ & \text{URD}[\text{nofURD}] = (\text{N}[\text{nofURD}], \text{R}[\text{nofURD}], \\ & \quad \text{nofKs}[\text{nofURD}], \text{Ks}[\text{nofURD}], \\ & \quad \text{D}[\text{nofURD}]) \end{aligned}$$

It fears the account of \*\*.

[0038] The representation information list unit 53 calculates the degree of similar between unread data memorized by the unread data-storage section 10, and consists of compression unread data read-out control-section 82 grade which controls read-out of unread data using the compression unread data-storage section 81 which memorizes the result of the unread data compression section 70 which decides which unread data are made all together, and the aforementioned unread data compression section 70, and compression unread data. The composition of the unread data compression section 70 is shown in drawing 13. 71 is the degree computing-control section of similar, 72 is the unread data read-out section, 73 is the degree vector generation section of similar, 74 is a switch, 75 is the reference vector storage section, 76 is the degree calculation section of similar, 77 is the threshold storage section, 78 is a comparator, 79 is a compression unread data write-in control section, 80 is the flag storage section, and 81 is the compression unread data-storage section.

[0039] The unread data output control section 11 receives a control signal DO, and the interface unit 51 reads the unread data signal URD [1] from the unread data-storage section 10 through the compression unread data read-out control section 82, and outputs a status signal DD. Moreover, the teacher signal T which shows whether the status signal DD is required for a user is received from a user, and the aforementioned teacher signal T, and the number signal nofKs of keywords [1] and the keyword group signal  $K_s$  [1] of the aforementioned unread data signal URD [1] are written in the teacher data-storage section 13 according to a predetermined procedure. The teacher data-storage section 13 is the teacher data signal of a maximum nofTD individual.

[Equation 3]

$TD[1]=(T[1],TnofK[1],TKs[1])$   
 $\vdots$   
 $TD[nofTD]=(T[noTD],TnofKs[nofTD],$   
 $TKs[nofTD])$

It fears the account of \*\*.

[0040] The study unit 52 consists of a portion which performs metric study which corrects affirmation/negative metric signal using the teacher signal T inputted by the user, and a portion which corrects the parameter for calculating a need signal from affirmation/negative signal, and a judgment parameter signal, and each portion is controlled by the learning control section.

[0041] In drawing 1, the composition of the portion which performs metric study is as follows.

[0042] The negative metric signal MN memorized by the affirmation metric signal MY and the aforementioned negative metric storage section 6 which were memorized by the affirmation metric storage section 5 in the metric study section 19 is corrected. This metric study section 19 reads the aforementioned teacher data TD from the teacher data-storage section 13, changes into a vector two or more keywords in the vector generation section 20 for study which is the same function as the vector generation section 1 of the study unit 50, is calculating an autocorrelation matrix and corrects affirmation/negative metric signal.

[0043] The composition of the portion which learns a judgment parameter signal is as follows.

[0044] The score calculation section 22 for study consists of the affirmation signal calculation section 221 for study, and the negative signal calculation section 222 for study. The affirmation signal calculation section 221 for study receives the vector signal for study from the vector generation section 20 for study, calculates the affirmation signal LSY for study, and the negative signal calculation section 222 for study receives the vector signal for study from the vector generation section 20 for study, and it calculates the negative signal LSN for study in it. The judgment side study section 21 rewrites the judgment parameter signal of the judgment parameter storage section 8 by the predetermined method in response to the judgment parameter learning control signal PLC from the learning control section 14. 14 controls switches 16, 17, and 18, the metric study section 19, the vector generation section 20 for study, the score calculation section 22 for study, the negative signal calculation section 23 for study, and the judgment side study section 21 by the learning control section in response to the study start signal LS.

[0045] It explains that rough processing with the information filter equipment constituted as mentioned above flows. In the following explanation, the history what "information" the user already needed in the past shall be study ending. Moreover, one or more keywords corresponding to "information" shall be given to what is only called "information." A keyword may be a part or the whole of each word which constitutes the "information" concerned, and since the "information" concerned is represented, you may attach it specially.

[0046] Record what "information" the user needed in the past will be read from the storage sections 2, 5, 6, and 8, and the information filtering unit 50 will estimate quantitatively the need for new "information" for the user concerned as a need signal, if new "information" is inputted.

[0047] next, the evaluated new "information" is unread from the past in the unread data-storage section 10 -- it is written in so that need signals including "information" may be located in a line with descending

[0048] In the representation information list unit 53, after the writing to the unread data-storage section 10 is completed, the unread data compression section 70 calculates the mutual degree of similar of the "information" memorized by the unread data-storage section 10, and it writes in the compression unread data-storage section 81 by making into a compression unread data signal the result which information is made all together.

[0049] an interface unit 51 -- a user -- descending of a need signal -- arranging -- the above -- new "information" was included -- unread -- typical information is shown when there are "information" and analogous information (for example, display display)

[0050] under the present circumstances, the above which was alike and the user was shown -- new "information" was included -- unread -- when a user inputs the teacher signal with which every \*\* of "information" shows [ the need or ] for a user whether it is unnecessary through an interface unit 51, in an interface unit 51, the teacher signal concerned is received and "information" and its teacher signal concerned are sent to the study unit 52

[0051] In addition, the input of the teacher signal by this user is carried out in order to heighten the machine learning of the study unit 52 more, and if the machine learning (machine learning of the history what "information" the user needed in the past) of the study unit 52 is already high fully, it is not necessary to perform it.

[0052] Next, in the study unit 52, the contents of a history of the storage sections 2, 5, 6, and 8 are rewritten using the "information" which carried out [ aforementioned ] presentation, and its teacher signal.

[0053] As mentioned above, the information filter equipment of this invention is adapted for a user through higher study, and can present preferentially the "information" which a user searches for with high list nature.

[0054] Moreover, in the initial state which is not learning, since a user does not understand what "information" is needed in the study unit 52 Although the input of the teacher signal by the user who mentioned above all the new "information" inputted whenever the user received presentation by the interface unit 51 is required It can be adapted for a user soon through the study carried out at any time, and the "information" which a user searches for can be preferentially shown now.

[0055] in addition, showing preferentially the "information" which a user searches for If the more concrete example of use describes, though the population A of a certain "information" database will be searched with a specific keyword and the "informational" reference set B will be obtained Even if not all the "information" on the reference set B concerned is necessarily required for a user and informational [ "informational" / all ] is required for a user, the required ranking is premised on naturally existing. Therefore, showing a user in order by the interface unit 51 according to needlessness or its required ranking in order of a required shell means showing preferentially the "information" which a user searches for.

[0056] Hereafter, the activity for every unit is concretely explained using a drawing.

[0057] An "informational" need signal is calculated in the information filtering unit 50.

[0058] Here, the fundamental concept for calculating a need signal is explained.

[0059] The case where the keyword is appended to the inputted "information" is considered. The keyword set C which is attached to the "information" made unnecessary with the keyword set A considering a one user attached to the "information" which the user needs by high frequency or high probability still better in all with the keyword set B attached by high frequency or high probability, or is not attached to it can be considered.

[0060] A value 0 is assigned to the keyword which belongs a negative value to the aforementioned keyword set C at the keyword which belongs a positive-number value to the aforementioned keyword set B at the keyword belonging to the aforementioned keyword set A, respectively.

[0061] And it judges to which keyword group of the aforementioned keyword sets A, B, and C each belongs about one or more keywords currently attached to the "information" newly inputted, and it constitutes so that the value assigned the account of before may be integrated. <BR> [0062] thus -- if constituted -- the above -- two or more keywords attached to the "information" newly inputted A positive big value is shown to the "information" (information that possibility that a user needs is high) in which many keywords belonging to the keyword set A were contained. It is convertible for the numeric value which shows a negative big value to the "information" (information that possibility of supposing that a user is unnecessary is high) to which many keywords belonging to the keyword set B are attached.

[0063] In this way, the need for a user can be predicted using the aforementioned numeric value. assignment of the value from required/unnecessary evaluation of the user about the "information" shown in this invention and its "information" to a keyword -- automatic -- carrying out -- calculation of a need signal with a high precision -- realizing - precision -- it has realized putting in order and changing "information" into the high, high order of need

[0064] Therefore, two or more keywords attached to "information" were changed into one vector, and the autocorrelation matrix of the aforementioned vector is separately calculated about the case where it is presupposed with the case where a user needs that it is unnecessary. the autocorrelation matrix MY made from the keyword currently attached to the "information" which the user answered the need -- using -- the length SY of Vector V -- [Equation 4]  

$$SY = \sum_i \sum_j M_{ij} \cdot V_i \cdot V_j$$

It calculates.

[0065] In addition, the autocorrelation matrix MN made from the keyword which attached the autocorrelation matrix MY made from the keyword currently attached to the "information" which answered that it was required to an "affirmation metric signal" and the information which answered that it was unnecessary is called "negative metric signal", and length SY is called affirmation signal. If many keywords well contained in the "information" which a user needs in two or more keywords which became the origin of Vector V are contained, the length SY of this length SY is effective, when calculating a need signal, since a positive big value is taken, and the value near 0 is taken when that is not right.

[0066] An example of the desirable initial state of information filter equipment is in the state which set altogether to -1 teacher signal [ of the minimum value Vmin and the teacher data TD of the teacher data-storage section 13 [j] with which the hardware which uses all need signal / of the unread data URD [i] of a zero matrix (nofDCKxnofDCK) and the unread data-storage section 10 / N [i] (i= 1, ..., nofURD) can express the affirmation metric signal MY and the negative metric signal It shall be such an initial state.

[0067] In the information filtering unit 50, first, the information data D are inputted from the information data input terminal 100, the number signal nofKs of keywords showing the number of the keyword attached to information data from the number signal input terminal 101 of keywords is inputted, and keyword group signal Ks= (K [1], K [2], ..., K [nofKs]) which is two or more keywords from the keyword signal input terminal 102 is inputted.

[0068] The keyword group signal Ks is changed into the vector signal V from the meeting of a character string by the vector generation section 1. By this conversion, the similarity of a keyword group signal can be calculated now as a distance of a vector.

[0069] It explains referring to the flow chart which shows operation of the vector generation section 1 to drawing 3 . First, if the number signal nofKs of keywords and the keyword group signal Ks are received ( drawing 3 step (\*\*)), internal vector signal V= (V [1], V [2], ..., V [nofDic]) will be set to (0, 0, ..., 0), and the keyword counter signal i will be set to 1 (this drawing step (\*\*), (\*\*)). Next, after setting the dictionary counter signal j zero times, only 1 increases the dictionary counter signal j (this drawing step (\*\*)).

[0070] Next, the sign dictionary signal DCK [j] which consists of a keyword specified by the dictionary counter j and a number is read from the dictionary storage section 2 which has the sign dictionary signal DCK of a nofDCK individual to the interior, and character string partial [ of the sign dictionary signal DCK ] W [j] and i-th keyword signal K [i] is compared (this drawing step (\*\*)). When both are not equal, only 1 increases the dictionary counter j (this drawing step (\*\*)). Processing of the drawing 3 step (\*\*) - (\*\*) is repeated until it becomes equal to the number nofDiC of the sign dictionary signal with which both were in agreement with the signal or the value of the dictionary counter j was stored in the dictionary storage section 2 (this drawing step (\*\*)).

[0071] If W [j] equal to keyword signal K [i] is found, j-th component V [j] of a vector signal will be set to 1 (this drawing step (\*\*)), and only 1 will increase the keyword counter signal i (this drawing step (\*\*)). Hereafter, same processing is performed until the keyword counter signal i becomes larger than the number signal nofKs of keywords (this drawing step (\*\*)).

[0072] In this way, in the vector generation section 1, the keyword group signal Ks which is the aggregate of the keyword signal which consists of a character string signal is changed into the vector signal V with the component of the nofDCK individual coded by 0 and 1.

[0073] Next, the affirmation signal calculation section 31 calculates the affirmation signal SY used as a big value, when many keywords contained in the information which the user needed for the keyword group signal Ks in the past are contained. For this reason, in response to the aforementioned vector signal V, the affirmation signal calculation section 31 reads the affirmation metric signal MY from the affirmation metric storage section 5, and is [Equation 5] about the affirmation signal SY.

$$SY = \sum_{i=0}^{nofDic-1} \sum_{j=0}^{nofDic-1} MY[i][j] \cdot V[i] \cdot V[j]$$

It calculates.

[0074] The negative signal calculation section 32 calculates the negative signal SN used as a big value, when many keywords contained in the information made unnecessary [ a user ] in the past are contained in the keyword group signal Ks. For this reason, the negative signal calculation section 32 reads the negative metric signal MN from the negative metric storage section 6, and is [Equation 6] about the negative signal SN.

$$SN = \sum_{i=0}^{nofDic-1} \sum_{j=0}^{nofDic-1} MN[i][j] \cdot V[i] \cdot V[j]$$

It calculates.

[0075] It is decided based on a response of the keyword group signal Ks and a user that the affirmation metric signal MY and the negative metric signal MN are mentioned later.

[0076] Thus, the information data D can be made to correspond to one on the two-dimensional space which took the affirmation signal SY along the vertical axis, and took the negative signal SN along the horizontal axis using the affirmation signal SY and the negative signal SN which were calculated, as shown in drawing 9 . That (it displays by O) for which a user needs a distribution of the information data D in this two-dimensional space is mainly distributed over the upper left section, and what is made unnecessary [ a user ] (it displays by x) mainly comes to be distributed over the lower right section.

[0077] Therefore, the information data D which a user needs, and the unnecessary information data D are separable by defining a suitable coefficient C, as shown in drawing 10 .

[0078] Furthermore, the need [ of being calculated using this coefficient C described below ] signal N serves as as big a



value as the information data D predicted that need is high, so that it is in the upper left in above-mentioned two-dimensional space. Therefore, if the information data D are put in order and shown to descending of the need signal N, a user can get required information efficiently. The reliability signal R of the direction which intersects perpendicularly with the need signal N is a signal which shows roughly what keyword signal was included in the dictionary among the keywords contained in the keyword group signal Ks. Therefore, the size of this reliability signal R shows which the need signal N which the information filter calculated can trust.

[0079] Next, the need calculation section 7 receives the aforementioned affirmation signal SY outputted from the aforementioned affirmation signal calculation section 31, and the aforementioned negative signal SN outputted from the aforementioned negative signal calculation section 32. A large number [ the keyword which read the judgment parameter signal C from the judgment parameter storage section 8, and was attached to the information which was the past need ] When there is almost no keyword currently attached to the unnecessary information, the need [ of becoming a big value ] signal N is calculated with  $N=SY-C-SN$ , and the reliability signal R is calculated with  $R=C-SY+SN$ .

[0080] It explains referring to the flow chart which showed operation of the unread data write-in control section 9 to drawing 4. First, it receives from each input terminal with the aforementioned information data D, and the aforementioned number signal nofKs of keywords and the aforementioned keyword group signal Ks, the aforementioned need signal N and the aforementioned reliability signal R are received from the need calculation section 7, and the unread data-processing signal WI outputted from the unread data-division directions terminal 110 is changed into 1 from 0 ( drawing 4 step (\*\*)). Next, it is referred to as  $i=1$  (this drawing step (\*\*)), and need signal [ of the unread data URD [i] memorized by the unread data-storage section 10 ]  $N[i]$  ( $i=1, \dots, \text{nofURD}$ ) is read one by one. As compared with the aforementioned need signal N (this drawing step (\*\*)), the number  $i1$  of the first unread data with which the aforementioned need signal N becomes large from need signal [ of the unread data URD [i] ]  $N[i]$  ( $N \geq N[i]$ ) is detected (this drawing step (\*\*) (\*\*)).

[0081] It is  $URD[i+1]=URD$  about the unread data 1st after  $i[i]$ . It replaces with  $i=i1, \dots, \text{nofURD}$  (this drawing step (\*\*) - (\*\*)). Then, it is N about the  $i1$  position unread data  $URD[i1]$  [ $i1$ ].  $=NR[i1]=R\text{nofKs}[i1]=\text{nofKsKs}[i1]=KsD[i1]$  It replaces by  $=D$ , the aforementioned need signal N, etc. (this drawing step (\*\*)). After this replacement is completed, the unread data-division indication signal WI outputted from the unread data-division directions terminal 110 is returned to 0 (this drawing step (\*\*)), and processing is ended.

[0082] Next, in the representation information list unit 53, if the degree calculation start signal of similar is inputted into the unread data-storage section 70 from a terminal 170, the degree computing-control section 71 of similar will set the internal information counter n and the condensed-information counter a to 1, respectively ( drawing 14 step (\*\*)).

[0083] The value of the counter output n of the aforementioned information counter and the counter output a of a condensed-information counter is sent to the compression unread data write-in control section 79. The compression unread data write-in control section 79 substitutes the value of the aforementioned information counter n for the similar information list signal AD [a] and [1]. Substitute the value 1 which shows that there is no analogous information in flag A [a], and 1 is substituted for the counter output l of the number-of-times counter of comparison. What added 1 at the value of the aforementioned information counter is substituted for the counter output m of 0 and a comparison information number counter at information number difference counter for comparison  $\text{deltam}$  (this drawing step (\*\*)).

[0084] The degree computing-control section 71 of similar sends the value of the information counter n to the unread data read-out section 72, and it changes it so that the reference vector storage section 75 may connect a switch 74. From the unread data-storage section 10, the unread data read-out section 72 reads the unread data  $URD[n]$  (this drawing step (\*\*)), and sends them to the degree vector generation section 73 of similar.

[0085] The degree vector generation section 73 of similar generates the degree vector signal of similar like the aforementioned vector generation section 1 (this drawing step (\*\*)), and sends it to the reference vector storage section 75. The reference vector storage section 75 memorizes the aforementioned degree vector signal of similar as a reference vector signal  $Vn$ . The degree calculation section 76 of similar reads the reference vector storage 75 reference vector signal  $Vn$ , and is [Equation 7].

$$\sum Vn[i]$$

It \*\*\*\*\* (this drawing step (\*\*)).

[0086] The degree computing-control section 71 of similar changes a switch 74 so that it may be connected with the degree calculation section 76 of similar, it reads the value of the counter output m of a comparison information number counter from the compressed data write-in control section 79, and sends it to the unread data read-out section 72.

[0087] The unread data read-out section 72 reads the unread data  $URD[m]$  from the unread data-storage section 10 (this drawing step (passing)), and sends the unread data  $URD[m]$  to the degree vector generation section 73 of similar.



[0088] The degree vector generation section 73 of similar generates the degree vector signal of similar like the aforementioned vector generation section 1 (this drawing step (\*\*)), and sends it to the degree calculation section 76 of similar as a comparison vector signal  $V_m$ .

[0089] The degree calculation section 76 of similar is [Equation 8] from the comparison vector signal  $V_m$ .

$$\sum_i V_m[i]$$

[Equation 9]

$$\sum_i V_n[i] \cdot V_m[i]$$

\*\* is calculated (this drawing step (\*\*), (\*\*)). The degree calculation section 76 of similar is [Equation 10].

$$\sum_i V_m[i]$$

\*\* -- [Equation 11]

$$\sum_j V_n[j]$$

When larger, it is [Equation 12] about the degree  $Ch$  of similar.

$$Ch = \frac{\sum_i V_n[i] \cdot V_m[i]}{\sum_i V_n[i]}$$

It carries out and, in the case of others, is [Equation 13].

$$Ch = \frac{\sum_i V_n[i] \cdot V_m[i]}{\sum_j V_m[j]}$$

It carries out. This degree  $Ch$  of similar is set to 1 when all the keywords attached to the unread data URD [n] are contained in the keyword attached to the unread data URD [m] (or when all the keywords attached to the unread data URD [m] are contained in the keyword attached to the unread data URD [n]), and in the case of others, it has the property to become small from 1.

[0090] Below, the property of this degree of similar is described as compared with the angle of the vector often used conventionally.

[0091] The angle of the vector used well has the property in which the case ( drawing 17 (b)) where it has the portion with which the case ( drawing 17 (a)) where one side includes another side between two keyword sets, and two sets lap only in a case by two vectors carrying out [ completeness ] although it is effective when expressing whether a difference is in the keyword which was not set to 0 but was attached is undistinguishable. For this reason, it is not suitable when expressing the inclusion relation of a keyword.

[0092] If the case where there are detailed information and rough information is actually considered about a certain incident and matter, the keyword by which the set of the keyword attached to detailed information was attached to rough information will often be seen included.

[0093] Therefore, when information thinks the degree of similar, the upper degree  $Ch$  of similar is more effective than the angle of the vector often used conventionally, in order to evaluate the informational degree of similar.

[0094] Now, a comparator 78 compares with threshold  $\theta_{ch}$  which had written beforehand the degree  $Ch$  of similar obtained in this way in the threshold storage section 77 ( drawing 15 step (\*\*)).

[0095] When the degree  $Ch$  of similar is larger than threshold  $\theta_{ch}$ , the compression unread data write-in control section 79 considers as 2 unread data URD [n], and URD [m] considers as the information on the same contents, and let the unread data URD [m] be the subordinate information on the unread data URD [n]. For this reason, it is made the value 1 showing being subordinate information about the flag signal  $f_{flag}$  [m], only 1 increases the value of condensed-information counter  $A$  [a], and the value of the comparison information number counter  $m$  is written in the similar information list signal  $AD$  [a] and  $A[a]$  (this drawing step (mosquito)).

[0096] When the degree  $Ch$  of similar is smaller than threshold  $\theta_{ch}$ , 2 unread data URD [n] and URD [m] consider as the information on different contents, and make the unread data URD [m] the independent information on the unread data URD [n]. For this reason, it is made the value 0 showing being independent information about the flag signal  $f_{flag}$  [m] (this drawing step (\*\*)).

[0097] Next, 1 is added to information number difference counter for comparison  $\Delta m$  (this drawing step (\*\*)).

[0098] Value  $\Delta m$  of the information number difference counter for comparison is added to the comparison information number counter  $m$ , and it compares with the number  $\text{nofURD}$  of unread data (this drawing step (\*\*)). If  $(m + \Delta m)$  is  $\text{nofURD}$ , there will be unread data which should still be compared. Then, it investigates whether the unread data  $\text{URD}[m + \Delta m]$  are independent data (this drawing step (\*\*)). That is, already, about subordinate information and the information which became, I hear that processing made into subordinate information in piles is not performed, and it is.

[0099] In not being independent information, 1 is added to information number difference counter for comparison  $\Delta m$  (this drawing step (\*\*)), and it performs same processing until independent information is found or unread data are lost. If independent information is found, only 1 will increase the value of the number-of-times counter  $l$  of comparison, and  $(m + \Delta m)$  will be substituted for the comparison information number counter  $m$  (this drawing step (\*\*)). If the value of the number-of-times counter  $l$  of comparison becomes below the number of times  $l_{\max}$  of the maximum comparison set up beforehand, the unread data  $\text{URD}[m]$  will newly be read and same processing will be performed (this drawing step (\*\*)).

[0100] In this way, the list of subordinate information made into the information on representation [ the unread data  $\text{URD}[n]$  ] is made as the similar information list signal  $\text{AD}[a]$  and  $[A[a]]$ .

[0101] the number of times  $l_{\max}$  of the maximum comparison which the value of the number-of-times counter  $l$  of comparison set up beforehand -- large -- becoming (this drawing step (\*\*)) -- the procedure which looks for the subordinate information made into the information on representation [ new unread data ] starts For that purpose, the unread data which are not processed as subordinate information are looked for like a top (this drawing step (\*\*), (\*\*), (\*\*), (\*\*)).

[0102] If the unread data which are not processed as subordinate information are newly found, only 1 will increase the value of  $(n + \Delta n)$  and the condensed-information counter  $a$  for the value of the information counter  $n$ , it will return to a step (b), and the above-mentioned processing will be performed.

[0103] This processing is repeated until the information counter  $n$  becomes more than the number  $\text{nofURD}$  of unread data.

[0104] The similar information list signal  $\text{AD}$  acquired by this processing becomes a thing like drawing 16 . The compression unread data write-in section 79 is written in the compression unread data-storage section 81 by making the value of the degree information signal  $\text{AD}$  of similar, and a condensed-information counter, i.e., the number of independent information, into compression unread data and the number  $\text{nofa}$  of compression unread data.

[0105] next, the unread data  $\text{UDR}$  are read with a compression unread data read-out control section, a user's response (teacher signal  $T$ ) is added, and teacher data signal  $\text{TD}$  is built -- it attaches interface unit 51 and explains

[0106] Next, it explains, referring to the flow chart which showed operation of the interface unit 51 to drawing 5 .

[0107] From the data read-out start signal input terminal 103, the data read-out start signal  $\text{DO}$  is inputted ( drawing 5 step (\*\*)). The unread data output control section 11 leads the compression unread data control section 82. The unread data  $\text{URD}[\text{AD}[1]]$  which are the 1st independent information, and  $[1]$  are read from the aforementioned unread data-storage section 10 (this drawing step (\*\*)). In being larger than the minimum value  $V_{\min}$ , need signal [ of unread data ]  $N[1]$  outputs to the data display terminal 104 by making information signal [ of the unread data signal  $\text{URD}[1]$  ]  $D[1]$  into the display information signal  $\text{DD}$ , and stands by (this drawing step (\*\*), (\*\*)). In being equal to the minimum value  $V_{\min}$ , need signal [ of unread data ]  $N[1]$  makes the display information signal  $\text{DD}$  "with no data", outputs to the data display terminal 104, and stands by (this drawing step (\*\*)).

[0108] In not being required, when a user (not shown) looks at the display information signal  $\text{DD}$  displayed on data display equipment (not shown), ending the teacher signal  $T = 1$  when it is required information, and ending the teacher signal  $T = 0$  and processing, it returns to the teacher signal input terminal 105 as teacher signal  $T = -1$  (this drawing step (\*\*)). In the case of teacher signal  $T = -1$ , processing is ended. in the case of teacher signal  $T \neq -1$ , (this drawing step and the unread data output control section 11 It is  $\text{TD}[i] = \text{TD}$  about the teacher data expressed with (several 2) of the teacher data-storage section 13  $[i-1]$ . It replaces with  $i = 2, \dots, \text{nofTD}$  (this drawing step (\*\*)). The aforementioned teacher signal  $T$ , and the number signal  $\text{nofKs}$  of keywords  $[1]$  and the keyword group signal  $\text{Ks}[1]$  of the aforementioned unread data are used for the 1st teacher data  $\text{TD}[1]$ , and it is  $T[1] = T \text{nofKs}[1] = \text{nofKs}[1]$   $\text{TKs}[1] = \text{Ks}[1]$

$\text{URD}[\text{AD}[a] \text{ and } [b]] = \text{URD}[\text{AD}[a+1], [b]]$ ,  $a = 1, \dots, (\text{nofa}-1)$ , it replaces (this drawing step (\*\*), (\*\*)). the unread data  $\text{URD}$  of the aforementioned unread data-storage section 10  $b = -1, 2$ , and -- carrying out (this drawing step (\*\*), (mosquito)) --  $\text{nofa}$  position independent unread data  $\text{URD}[\text{-- the need signal of } \text{AD}[\text{nofa}] \text{ and } [1] \text{ -- } N[\text{nofa}] = (\text{minimum value } V_{\min})$

It carries out (this drawing step (\*\*), (\*\*), (\*\*)).

[0109] Next, it explains, referring to the flow chart shown in drawing 6 - drawing 8 about operation of the study unit 52.

[0110] the flow chart example which shows the outline of operation of the learning control section 14 to drawing 6 -- it explains in detail

[0111] In drawing 6, first, the study start signal LS is inputted from the study start signal input terminal 106, the learning control section indication signal LI outputted from the learning control section indication signal output terminal 107 is changed into 1 from 0 (drawing 6 step (\*\*)), and under processing is shown. Next, it changes so that the metric study section 19 and the vector generation section 20 for study may connect a switch 16, a switch 17, and a switch 18 (this drawing step (\*\*)).

[0112] Next, the metric study section 19 corresponding to step (\*\*) of drawing 7 is operated (after operating this drawing step (\*\*)) and the judgment side study section 21 (this drawing step (\*\*)), processing is ended using LI as 0 (this drawing step (\*\*)).

[0113] Next, the metric study section 19 explains operation which corrects affirmation/negative metric signal using drawing 7 using a user's response (teacher signal T) and the keyword group signal Ks.

[0114] Drawing 7 is the flow chart of operation of the metric study section 19, and the metric (drawing 7 step (\*\*)) study section 19 which received the metric learning control signal MLC from the aforementioned learning control section 14 reads the negative metric storage section 6 to the negative metric signal MN for the affirmation metric signal MY from the affirmation metric storage section 5 in this drawing, respectively.

[0115] Next, the metric study section 19 sets the value of the teacher data counter c to 1 (this drawing step (\*\*)). Next, c-th teacher data signal TD [c] is read from the teacher data-storage section 13 (this drawing step (\*\*)), and teacher signal [ of the teacher data TD [c] ] T [c] is investigated. When aforementioned teacher signal T [c] is not -1 ( $T \neq -1$ ), (this drawing step (\*\*)), and the number signal TnofKs of keywords [c] and the keyword group signal TKs [c] of the teacher data TD [c] are outputted (this drawing step (\*\*)). The vector generation section 20 for study which received the number signal TnofKs of keywords [c] and the keyword group TKs [c] of the aforementioned teacher data TD [c] performs the same operation as the vector generation section 1 of the above-mentioned information filtering unit 50, and outputs the vector signal LV for study (this drawing step (\*\*)). The metric study section 19 is  $MY[i][j] = MY[i][j] + LV[i][j]$  and LV [j] about (this drawing step (\*\*)) and the affirmation metric signal MY, when the aforementioned vector signal LV for study is received and teacher signal [ of the aforementioned teacher data TD [c] ] T [c] is  $T = 1$ .  
(-- here, it corrects with i, and  $j = 1 - \text{nofDiC}$ ) (this drawing step (\*\*))

[0116] By this processing, the affirmation metric signal MY comes to have a big value to the keyword signal (plurality) currently attached to the information data D which the user needed. Consequently, the above-mentioned affirmation signal SY comes to become large to the information data D which a user needs. The processing same as follows is made also for the negative metric signal MN.

[0117] the case where teacher signal [ of the aforementioned teacher data TD [c] ] T [c] is  $T = 0$  -- the negative metric signal MN --  $MN[i][j] = MN[i][j] + LV[i][j]$  and LV [j]

(-- here, it corrects with i, and  $j = 1 - \text{nofDiC}$ ) (this drawing step (\*\*))

[0118] Only 1 increases the value of a teacher data counter as  $c = c + 1$  (this drawing step (\*\*)).

[0119] Hereafter, the metric study section 19 repeats the same operation until teacher signal [ of the teacher data TD [c] ] T [c] is set to  $T[c] = -1$  or it serves as  $c = \text{nofTD}$ . When it comes to  $T[c] = -1$  or  $c = \text{nofTD}$ , processing of (this drawing step (\*\*)) and metric study is ended, and the metric learning control signal MLC is sent to the learning control section 14.

[0120] In response to the metric learning control signal MLC from the metric study section 19, the learning control section 14 is changed so that the vector generation section 20 for study and the score calculation section 22 may connect a switch 16, and it is changed so that the vector generation section 20 for study and the judgment side study section 21 may connect a switch 17 and a switch 18. The learning control section 14 sends the judgment side learning control signal PLC to the judgment side study section 21.

[0121] Next, the judgment side study section 21 is explained in detail using drawing 8.

[0122] The judgment side study section 21 asks for the coefficient C which separates best the information data D made unnecessary [ the information data D and the user whom the user expressed on two-dimensional space using the affirmation signal SY and the negative signal SN needs ], as shown in drawing 10.

[0123] According to the flow chart which showed processing for this to drawing 8, it explains in detail.

[0124] First, the value of the teacher data counter c is set to 1 in response to the aforementioned judgment side learning control signal PLC (drawing 8 step (\*\*)) (this drawing step (\*\*)). c-th teacher data signal TD [c] is read from the teacher data-storage section 13 (this drawing step (\*\*)), and teacher signal [ of the teacher data TD [c] ] T [c] is investigated (this drawing step (\*\*)). When aforementioned teacher signal T [c] is not -1 ( $T \neq -1$ ), the number signal

TnofKs of keywords [c] and the keyword group signal TKs [c] of the teacher data TD [c] are outputted (this drawing step (\*\*)). The vector generation section 20 for study which received the number signal TnofKs of keywords [c] and the keyword group TKs [c] of the aforementioned teacher data TD [c] performs the same operation as the vector generation section 1 of the information filtering unit 50 mentioned above, and outputs the vector signal LV for study.

[0125] The score calculation section 22 for study performs the same operation as the score calculation section 3 of the information filtering unit 50 mentioned above, the affirmation signal LSY for study [c] and the negative signal LSN for study [c] are outputted, and the judgment side study section 21 receives it (this drawing step (\*\*)). The aforementioned affirmation signal LSY for study [c], the aforementioned negative signal LSN for study [c], teacher signal [ of the teacher data TD [c] ] T [c], and signal TC for judgment side study [c] = (T [c], LSN [c], LSY [c]) are memorized to an internal storage element (this drawing step (\*\*)). And only 1 increases the value of a teacher data counter as  $c=c+1$  (this drawing step (\*\*)).

[0126] Hereafter, the judgment side study section 21 repeats the same operation until teacher signal [ of the teacher data TD [c] ] T [c] is set to T[c] = -1 or it serves as  $c=\text{nofTD}+1$  (this drawing step (\*\*)). When it comes to T[c] = -1 or  $c=\text{nofTD}$ , processing of the affirmation signal LSY [c] calculation for study etc. is ended.

[0127] Next, the signal TC for judgment side study [c] ( $c=1, \dots$ ) with which the judgment side study section 21 was memorized by the internal storage element will serve as a distribution as shown in drawing 9, if a horizontal axis is set to LSN [c] and it sets a vertical axis to LSY [c], and O shows T[c] = 1 and it shows T[c] = 0 by x. What is teacher signal T[c] = 1, and the thing which is aforementioned teacher signal T[c] = 0 calculate the judgment parameter C which can be best separated as shown in drawing 10 by the climbing-a-mountain method among these (this drawing step (\*\*)). Next, the aforementioned judgment parameter C is written in the judgment parameter storage section 8, the judgment side learning control signal PLC is sent to the learning control section 14 (this drawing step (\*\*)), and processing is ended. The learning control section 14 receives the judgment side learning control signal PLC from the judgment side study section 21, makes it the value which shows under the standby by the learning control section indication signal, and ends processing.

[0128] As shown in drawing 10, unnecessary information mainly comes to be distributed over the upper left at the lower right by the information for which a user needs a keyword group signal on the two-dimensional space expressed with the affirmation signal SY and the negative signal SN using two above-mentioned metric signals. therefore, the above -- it comes to take a big value to the information for which, as for  $N=SY-C-SN$ , then a need signal, a user needs a need signal using a suitable coefficient C like

[0129] In addition, the cost function constituted as the calculation method of the judgment parameter C here based on the distance of a judgment side, the need signal LN for study, and the reliability signal LR for study although the climbing-a-mountain method was adopted [several 14]

$$\text{COST} = \sum_c (2 \cdot T[c] - 1) (LSN[c] - C \cdot LSY[c])$$

You may be the method of asking for the judgment side parameter C made into \*\*\*\*\* by Newton's method, the attacking method, etc.

[0130] moreover, the effect of oblivion of study of the affirmation metric signal MY and the negative metric signal MN was put in --  $MY[i][j] = \alpha \cdot MY[i][j] + LV[i]$  and  $LV[j]$

$$MN[i][j] = \beta \cdot MN[i][j] + LV[i] \text{ and } LV[j]$$

A result also with sufficient \*\*\*\*\* is obtained. (Positive number with beta smaller here than alpha and 1) the information to which the keyword is not given if the composition which adds the keyword generation section which generates a keyword group signal and the number signal of keywords further from the document indicated by reference "an Information Processing Society of Japan technical report and natural language processing 101-8 (1994. 5.27)" etc. is taken -- receiving -- \*\* -- applicable information filter equipment can be constituted

[0131] About the information to which the title was attached, it may consider as a keyword with the word which constitutes a title, and the number signal of keywords and a keyword group signal may be generated.

[0132] In addition, even if it also makes it a keyword signal contain class marks, such as the International Patent Classification number, it does not need to change the composition of this invention and can obtain a good result.

[0133] Moreover, it is easy to take composition which information filter equipment is correctly told that it is to which unread data with which two or more unread data URD were simultaneously displayed, and two or more users were displayed it answered although the form 1 of operation of this invention showed the case where every one unread data URD was shown depending on the size of display (not shown).

[0134] Moreover, although it constituted from a form 1 of operation of this invention so that only independent information including representation information might be displayed, the number or existence of subordinate

information is doubled and you may make it make it display by demand of a user. Moreover, you may express subordinate information as a small character and a picture in the range which does not spoil the list nature of independent information if needed.

[0135] When showing a user information in order of need since it was made to indicate representation information and the independent information by the list while according to the information filter equipment of the form 1 of operation calculating the degree of similar between information and putting similar information together, the fault that the information on the same contents is shown repeatedly can be prevented. Furthermore, although it is not the main purpose of this invention, it is also possible to use this invention as simple "equipment which puts similar information together."

[0136] According to the information filter equipment of the form 1 of operation, the relation of the response and keyword of a user is made to reflect in the affirmation metric signal MY which observed the simultaneous appearance of a keyword, and a negative metric signal, and the sign information of a keyword is projected on the space where distance was defined by changing a keyword group signal into the affirmation signal SY and the negative signal SN using these two metric signals. This can estimate the distance of a keyword group now with the analog scale of distance.

unnecessary [ in a Prior art / in the need ] by using this -- it becomes possible for evaluation of that need only whose alternative judgment was completed 2 persons to arrange in order of the need for a user

[0137] According to the information filter equipment of the form 1 of operation, to the information which a user needs, it comes to take a value with a big need signal, consequently what [ one ] has typical analogous information comes to be preferentially displayed for the information that need is high on interface units, such as display, for a user by the study based on the teacher signal from a user.

[0138] (Form 2 of operation) The form of operation of the 2nd of this invention is explained hereafter, referring to a drawing. The form 2 of operation is improved to what added the dictionary study section to the composition of the form 1 of operation, updated so that the sign dictionary signal DCK memorized by the dictionary storage section 2 might be adapted for a user, and took into consideration the probability distribution of the appearing keyword that information is required / unnecessary from the autocorrelation matrix of the keyword corresponding to a simple frequency distribution for the affirmation metric signal MY and the negative metric signal MN.

[0139] Although the block schematics of the information filter equipment of the form 2 of operation of this invention are shown in drawing 11 , different composition from the block schematics of the information filter equipment of the form 1 of operation of this invention mentioned above is explained in detail.

[0140] The dictionary study section which 23 receives the dictionary study signal DLC from the learning control section 14, and updates the sign dictionary signal DCK of the dictionary storage section 2 in drawing 11 , When, as for 24, the character string W and Number C are included in the keyword group signal Ks the table where a user consists of the number of times PN of negative which shows the number of times which the information data D answered are unnecessary when the number of times PY of affirmation and the character string W which show the number of times to which was resembled and the user answered the information data D are required are contained in the keyword group signal Ks -- a noFDCK piece -- the adaptation sign dictionary signal which it has

[Equation 15]

$$FDCK[1]=(W[1],C[1],PY[1],PN[1])$$

⋮

$$FDCK[nofFDCK]=(W[nofFDCK],C[nofFDCK], \\ PY[nofFDCK],PN[nofFDCK])$$

The account of \*\* The feared adaptation sign dictionary signal storage section, the number-of-times storage section which memorizes the number-of-times signal NN of total negation which shows the number of times which answered that it was unnecessary with all the number-of-times signals NY of affirmation that show the number of times which answered that a user was required for 25, The primary affirmation metric storage section 26 remembers the primary affirmation metric signal MY1 for renewal of affirmation metric to be, The primary negative metric storage section 27 remembers the primary negative metric signal MN1 for renewal of negative metric one to be, 28 calculates the affirmation metric signal MY and the negative metric signal MN which were improved from the aforementioned number-of-times signal of affirmation, the aforementioned number-of-times signal of negative, the aforementioned primary affirmation metric signal MY1, and the aforementioned primary negative metric signal MN1. It is KD metric study section which writes each in the affirmation metric storage section 5 and the negative metric storage section 6.

[0141] About the information filter equipment constituted as mentioned above, operation is explained using a drawing. However, the part as the form 1 of operation where operation is the same omits explanation.

[0142] An example of the desirable initial state of information filter equipment The affirmation metric signal MY and the negative metric signal MN A zero matrix (nofDCKxnofDCK), The minimum value Vmin which can express the hardware which uses all need signal [ of the unread data URD of the unread data-storage section 10 [i] ] N [i] (i= 1, ..., nofURD) Altogether the character string W of -1 and an adaptation sign dictionary signal for teacher signal [ of the teacher data TD of the teacher data-storage section 13 [j] ] T [j] A blank, It is in the state which corresponded Number C in order of the upper shell of the sign dictionary signal FDCK, corresponded 1, 2, ..., nofFDCK, the number of times PY of affirmation, and the number of times PN of negative to 0 and the adaptation sign dictionary, and also made all the character strings of a sign dictionary blank.

[0143] First, operation of the information filtering unit 50 is explained.

[0144] In the case of an above-mentioned initial state, the information filtering unit 50 performs operation as indicated in the form 1 of operation, both the need signal N and the reliability signal R are calculated with 0 from the number signal nofKs of keywords, the keyword group signal Ks, and the information data D which were inputted, and it stores in the unread data-storage section 10. Next, an interface unit 51 performs the same operation as the form 1 of operation, and sends the teacher data TD to which a user's response was attached to the teacher data-storage section 13.

[0145] Operation of the unread data compression section 70 is the same as that of the form 1 of operation.

[0146] As for operation of the study unit 52, the study start signal LS is first inputted from the study start signal input terminal 106. The learning control section 14 changes into 1 from 0 the learning control section indication signal LI outputted from the learning control section indication signal output terminal 107 in response to the aforementioned study start signal LS, and shows under processing. Furthermore, the dictionary study signal DLC is sent to the dictionary study section 23.

[0147] Operation of the dictionary study section 23 is explained referring to the flow chart shown in drawing 12 hereafter. First, in response to the dictionary study signal DLC (drawing 12 step (\*\*)), the adaptation sign dictionary FDCK is read into the adaptation code-signal buffer which can memorize the adaptation code signal of a maximum nofFDCKtmp individual from the adaptation sign dictionary storage section 24. The primary negative metric signal storage section 27 to the primary negative metric signal MN1 is read [ all the number-of-times signals NY of affirmation, and the number-of-times signal NN of total negation ] for the primary affirmation metric storage section 26 to the primary affirmation metric signal MY1 from the number-of-times storage section 25 (this drawing step (\*\*)). Next, the value of the internal teacher data counter c is set to 1 (this drawing step (\*\*)), the teacher data TD [c] are read from the teacher signal storage section 13 (this drawing step (\*\*)), and it judges whether teacher signal T [c] is -1 (this drawing step (\*\*)).

[0148] In the case of T[c] !=-1, the following processings are performed. First, the value of the internal number counter i of keywords is set to 1 (this drawing step (\*\*)), and the value of the adaptation sign dictionary counter j is set to 1 (this drawing step (\*\*)). Next, it judges whether aforementioned character string W [j] is blank or there is nothing (this drawing step (\*\*)), and in being blank, it replaces aforementioned character string W [j] by the aforementioned keyword signal TK [i] (this drawing step (\*\*)). In not being blank, it compares character string [ of the i-th keyword signal TK [i] of the teacher data TD [c], and the j-th adaptation sign dictionary signal FDCK [j] ] W [j] (this drawing step (\*\*)).

[0149] When aforementioned character string W [j] is blank, or when it is not blank and the aforementioned keyword signal TK [i] and aforementioned character string W [j] are in agreement, the following processings are performed according to the value of T [c]. In the case of T[c]=1 (this drawing step (\*\*)), 1 is added to all the affirmation signals NY (this drawing step (\*\*)), and 1 is added to the number of times PY of affirmation [j] of the adaptation sign dictionary signal FDCK [j] (this drawing step (\*\*)). Although T[c] !=1 and this are the cases of T[c]=0, 1 is added to the total negation signal NN (this drawing step (mosquito)), and 1 is added to the number of times PN of negative [j] of the adaptation sign dictionary signal FDCK [j] (this drawing step (\*\*)).

[0150] When the aforementioned W [j] is not blank and the aforementioned keyword signal TK [i] and aforementioned character string W [j] are not in agreement, the value of the adaptation sign dictionary counter j is increased one (this drawing step (\*\*)). The value of the adaptation sign dictionary counter j compares with value nofFDCKtmp+1 which added 1 to the number of adaptation code signals memorizable to an adaptation sign dictionary signal buffer (this drawing step (\*\*)). In the case of not more than nofFDCKtmp+1, the value of the adaptation sign dictionary counter j returns to the judgment with blank character string W [j].

[0151] When other, only 1 increases the value of the aforementioned keyword counter i (this drawing step (\*\*)).

[0152] When small as compared with value TnofKs+1 which added 1 to the number signal TnofKs of keywords of the aforementioned teacher data TD [c] (this drawing step (\*\*)), the value of the aforementioned keyword counter i sets the dictionary counter j to 1, and performs same processing. When other, only 1 increases the value of the teacher data counter c (this drawing step (\*\*)). The value of the teacher data counter c is compared with value nofTD+1 which added 1 to the number nofTD of teacher data (this drawing step (\*\*)), when the value of the teacher data counter c is small, the



following teacher data TD [c] are read and same processing is performed.

[0153] The above processing is performed to all the teacher data TD.

[0154] Next, the dictionary study section 23 calculates the keyword cost signal KD to each adaptation sign dictionary signal FDCK [j]. This keyword cost signal is an amount used in order to judge whether character string W [j] is effective as a keyword.

[0155] By the way, probability NN/in which a user's unnecessary information data D appear (NY+NN)

The information data D to which it compares and character string W [j] is attached are probability [ in the case of being unnecessary for a user ]  $PN[j]/(PY[j]+PN[j])$ .

\*\* -- if it is a thing so that it may become large when it differs greatly, character string W [j] is effective when the information data D judge with it being unnecessary for a user Probability NY/in which similarly a user's required information data D appear (NY+NN)

Probability [ when the information data D to which it compares and character string W [j] is attached are required for a user ]  $PY[j]/(PY[j]+PN[j])$

\*\* -- if it is a thing so that it may become large when it differs greatly, character string W [j] is effective when the information data D judge with it being required for a user

[0156] The keyword cost signal KD is [Equation 16] which will be called cull back divergence as one of the desirable examples although it is good anything if it is in the amount reflecting this property.

$$NY/(NY+NN) \cdot \log((PY[j])/(PY[j]+PN[j])) \\ + NN/(NY+NN) \cdot \log((PN[j])/(PY[j]+PN[j]))$$

\*\*\*\*\*. However, this has the cases of being unsuitable, such as overestimating the keyword cost signal of the adaptation sign dictionary signal FDCK [j] which fills  $PY[j]+PN[j] \cdot 1$  which cannot perform calculation of  $\log()$ , all the number-of-times signals NY of affirmation, such as an initial state of this information filter equipment, the number-of-times signal NN of total negation, the number of times PY of affirmation [j], and the number of times PN of negative [j] are 0, if it remains as it is. One of the desirable forms of operation which avoids this is [Equation 17] about a keyword cost signal.

$$KD[j] \\ = \tanh((PY[j]+PN[j])/PC) \cdot \\ \tanh\{NY/(NY+NN) \\ \cdot \log((PY[j]+\epsilon)/(PY[j]+PN[j]+2\epsilon)) \\ + NN/(NY+NN) \\ \cdot \log((PN[j]+\epsilon)/(PY[j]+PN[j]+2\epsilon))\}$$

It carries out. Here, epsilon is a parameter with the positive small value for avoiding the division of 0, and  $\log 0$ . As for Parameter PC, it is good to consider as about three to 20 value.

[0157] Next, character string [ of the adaptation sign dictionary signal FDCK [j] ] W [j], the number of times PY of affirmation [j], and the number of times PN of negative [j] are rearranged into descending of the keyword cost signal KD (this drawing step (\*\*)). At this time, the first order of a list remains in number [ of the adaptation sign dictionary FDCK [j] ] C [j]. This is used. From the primary affirmation metric signals MY1 and C [j] to C [i] the value of C [j] -- both -- several [ of the sign dictionary DCK ] -- case it is smaller than noFDCK --  $M[i][j] = MY1[ - C[i] ] [ - C[j] ]$  i, j= 1, and .. in the case of noFDCK and others the case of i=j --  $M[i][i] = PY[ - C[i] ]$  i= 1 and .. in noFDCK! = j  $M[i][j] = 0$  i, j= 1, and .. after being referred to as noFDCK --  $MY1[i][j] = M[i][j]$ , i, j= 1, and .. replacement of noFDCK and the primary affirmation metric signal MY1 is performed The same replacement is performed also to the primary negative metric signal MN1 (this drawing step (\*\*)).

[0158] And number [ of the adaptation sign dictionary FDCK in an adaptation sign dictionary signal buffer [j] ] C [j] is replaced with C[j] = j, j= 1, ..., noFDCKtmp.

[0159] After finishing the above processing, the dictionary study section 23 writes character string [ of the high order noFDCK individual of the adaptation sign dictionary FDCK in an adaptation sign dictionary buffer ] W [j], and number C [j] in the dictionary storage section 2. The high order noFDCK individual of the adaptation sign dictionary signal FDCK in an adaptation sign dictionary buffer [j] is written in the adaptation sign dictionary storage section 24. All the number-of-times signals NY of affirmation and the number-of-times signal NN of total negation are written in the number-of-times storage section 25, the primary affirmation metric signal MY1 is written in the primary affirmation metric signal storage section 26, and the primary negative metric signal MN1 is written in the primary negative metric signal storage section 27 (this drawing step (\*\*)).

[0160] Finally, the dictionary study signal DCL is returned to the learning control section 14 (this drawing step (\*\*)), and processing is ended.

[0161] Next, the aforementioned learning control section 14 is changed so that the metric study section 19 and the vector generation section 20 for study may connect a switch 16, a switch 17, and a switch 18. The aforementioned learning control section 14 sends the metric learning control signal MLC to KD metric study section 28.

[0162] KD metric study section 28 which received the aforementioned metric learning control signal MLC reads the primary negative metric storage section 27 to the primary negative metric signal MN1 for the primary affirmation metric signal MY1 from the primary affirmation metric storage section 26 first, respectively.

[0163] Next, KD metric study section 28 sets the value of the teacher data counter c to 1.

[0164] c-th teacher data signal TD [c] is read from the teacher data-storage section 13, and teacher signal [ of the teacher data TD [c] ] T [c] is investigated. When aforementioned teacher signal T [c] is not -1 (T!= -1), the number signal TnofKs of keywords [c] and the keyword group signal TKs [c] of the teacher data TD [c] are outputted. The vector generation section 20 for study which received the number signal TnofKs of keywords [c] and the keyword group TKs [c] of the aforementioned teacher data TD [c] performs the same operation as the vector generation section 1 of the information filtering unit 50 of the form 1 of operation mentioned above, and outputs the vector signal LV for study. the case where KD metric study section 28 receives the aforementioned vector signal LV for study, and teacher signal [ of the aforementioned teacher data TD [c] ] T [c] is T= 1 -- the primary affirmation metric signal MY1 -- MY1 -- [i] [j] =MY1[i] [j]+LV[i] and LV [j]

(-- here, it corrects with i, and j= 1 - nofDiC) the case where teacher signal [ of the aforementioned teacher data TD [c] ] T [c] is T= 0 -- the primary negative metric signal MN1 -- MN1 -- [i] [j] =MN1[i] [j]+LV[i] and LV [j]

(-- here, it corrects with i, and j= 1 - nofDiC) Only 1 increases the value of a teacher data counter as c=c +1.

[0165] Hereafter, KD metric study section 28 repeats the same operation until teacher signal [ of the teacher data TD [c] ] T [c] is set to T[c] =-1 or it serves as c=nofTD. When it comes to T[c] =-1 or c=nofTD, study of the primary affirmation metric signal MY1 and the primary negative metric signal MN1 is finished.

[0166] Next, all the number-of-times signals NY of affirmation and the number-of-times signal NN of total negation are read from the number-of-times storage section 25, and the affirmation metric signal MY is calculated using the primary affirmation metric signal MY1 and the primary negative metric signal MN1.

[0167] In this way, for the affirmation metric signal MY calculated and the negative metric signal MN, the affirmation signal SY and the negative signal SN which are calculated as well as the keyword cost signal KD are probability NN/(NY+NN) in which a user's unnecessary information data D appear.

The information data D to which it compares and character string W [j] is attached are probability [ in the case of being unnecessary for a user ] PN [j]/(PY[j]+PN [j]).

\*\* -- probability NY/(NY+NN) in which it is a thing so that it may become large, and a user's required information data D appear when it differs greatly

Probability [ when the information data D to which it compares and character string W [j] is attached are required for a user ] PY [j]/(PY[j]+PN [j])

\*\* -- if it has the property to be a thing so that it may become large when it differs greatly, it is good anything The desirable thing which fills this is [Equation 18] about the affirmation metric signal MY.

MY[i][j]

$$=NY/(NY+NN)$$

$$\cdot \log((MY1[i][j]+\epsilon) \cdot (NY+NN))$$

$$/(NY \cdot (MY1[i][j]+MN1[i][j]+2\epsilon))$$

It calculates and is [Equation 19] about the negative metric signal MN.

MN[i][j]

$$=NN/(NY+NN)$$

$$\cdot \log((MN1[i][j]+\epsilon) \cdot (NY+NN))$$

$$/(NN \cdot (MY1[i][j]+MN1[i][j]+2\epsilon))$$

It calculates. Here, epsilon is a parameter with the positive small value for avoiding the division of 0, and log0.

[0168] And the negative metric signal MN newly calculated [ signal / primary negative metric / MN1 / which was updated by the primary affirmation metric signal storage section 26 in the updated primary affirmation metric signal MY1 ] by the primary negative metric signal storage section 27 in the calculation \*\*\*\* affirmation metric signal MY to the affirmation metric storage section 5 is written in the negative metric storage section 6. Above, KD metric study



section 28 ends processing of metric study, and sends the metric learning control signal MLC to the learning control section 14.

[0169] In response to the metric learning control signal MLC from KD metric study section 28, the learning control section 14 is changed so that the vector generation section 20 for study and the score calculation section 22 may connect a switch 16, and it is changed so that the vector generation section 20 for study and the judgment side study section 21 may connect a switch 17 and a switch 18. The learning control section 14 sends the judgment side learning control signal PLC to the judgment side study section 21.

[0170] Since operation of the judgment side study section 21 is completely the same as the form 1 of operation, explanation is not repeated.

[0171] At once, since the sign dictionary of the dictionary storage section 2 is no longer empty when the above processing is performed, the need [ of being outputted from the information filtering unit 50 ] signal N, and the reliability signal R stop being 0, and the high information data of the need for a user come to be written in the high order of the unread data-storage section 10.

[0172] Henceforth, in order to judge whether it is the information which a user needs by repeating the above-mentioned processing, an effective keyword comes to be preferentially memorized by the dictionary storage section 2, and even if it is a small-scale dictionary, information filtering with a high precision is attained.

[0173] In addition, as the calculation method of the judgment parameter C, although the climbing-a-mountain method was adopted, you may be the method of asking for the judgment side parameter C which makes the maximum the cost function constituted based on the distance of a judgment side, the need signal LN for study, and the reliability signal LR for study by Newton's method, the attacking method, etc. like the form 1 of operation here. furthermore -- as a simple method --  $C = \tan \theta$  -- here --  $\theta = 0.5 \text{ and } \pi (i/90)$  -- the method referred to as choosing C which can separate best the information which is  $T[c] = 1$ , and the information which is  $T[c] = 0$  out of 1, ..., 90 can also be considered

[0174] moreover, the effect of oblivion of study of the primary affirmation metric signal MY1 and the primary negative metric signal MN1 was put in --  $MY1[i][j] = \alpha \cdot MY1[i][j] + LV[i][j]$  and  $LV[i][j]$

$MN1[i][j] = \alpha \cdot MN1[i][j] + LV[i][j]$  and  $LV[i][j]$

A result also with sufficient \*\*\*\*\* is obtained. (Here, alpha is a positive number smaller than 1) or When either  $MY1[i][j]$  or  $MN1[i][j]$  are able to come constant value, it is desirable operationally to constitute as  $MY1[i][j] = MY1[i][j] / 2$  and  $MN1[i][j] = MN1[i][j] / 2$ , so that overflow of a signal may be prevented. This is the same also about the number of times PY of affirmation [j], the number of times PN of negative [j], and all the number-of-times signals NY of affirmation and the number of times NN of total negation of the adaptation sign dictionary signal FDCK [j].

Furthermore, if the composition which adds the keyword generation section which generates a keyword group signal and the number signal of keywords from the document indicated by reference "an Information Processing Society of Japan technical report and natural language processing 101-8 (1994. 5.27)" etc. is taken, information filter equipment applicable also to the information to which the keyword is not given can be constituted.

[0175] About the information to which the title was attached, it may consider as a keyword with the word which constitutes a title, and the number signal of keywords and a keyword group signal may be generated.

[0176] In addition, even if it also makes it a keyword signal contain class marks, such as the International Patent Classification number, it does not need to change the composition of this invention and can obtain a good result.

[0177] Moreover, it is easy to take the composition which tells about which unread data URD two or more unread data URD were displayed simultaneously, and the user answered correctly to information filter equipment although the form of this operation showed the case where every one unread data URD was shown depending on the size of display (not shown).

[0178] Moreover, although it constituted from a form 1 of operation of this invention so that only independent information might be displayed, it is easy to constitute by demand of a user, so that subordinate information can also be displayed.

[0179] When showing a user information in order of need since it was made to indicate representation information and the independent information by the list while according to the information filter equipment of the form 2 of operation calculating the degree of similar between information and putting similar information together, the fault that the information on the same contents is shown repeatedly can be prevented.

[0180] According to the information filter equipment of the form 2 of operation, it is in having projected the sign information of a keyword on the space where distance was defined by introducing metric one which observed the simultaneous appearance of a keyword. This can estimate the distance of a keyword group now with the analog scale of distance. unnecessary [ in a Prior art / in the need ] by using this -- it becomes possible for evaluation of that need only whose alternative judgment was completed 2 persons to arrange in order of the need for a user

[0181] According to the information filter by the gestalt of this operation, to the information which a user needs, it

comes to take a value with a big need signal, consequently what [ one ] has typical analogous information comes to be preferentially displayed for the information that need is high on display for a user by the study based on the teacher signal from a user.

[0182]

[Effect of the Invention] As mentioned above, the storage section the information to which this invention was supplied from the information record medium or the information communication network is remembered to be, It is providing a similar judging means summarizing the information on this storage section for similar information, and deciding representation information and subordinate information, and a presentation means showing representation information about the information which is similar when showing the information on the aforementioned storage section. When two or more same information is included, by [ typical / of information and related information ] indicating by the number of cases, as many kinds as possible of information can be displayed at once, and high information filtering of list nature can be realized.

---

[Translation done.]

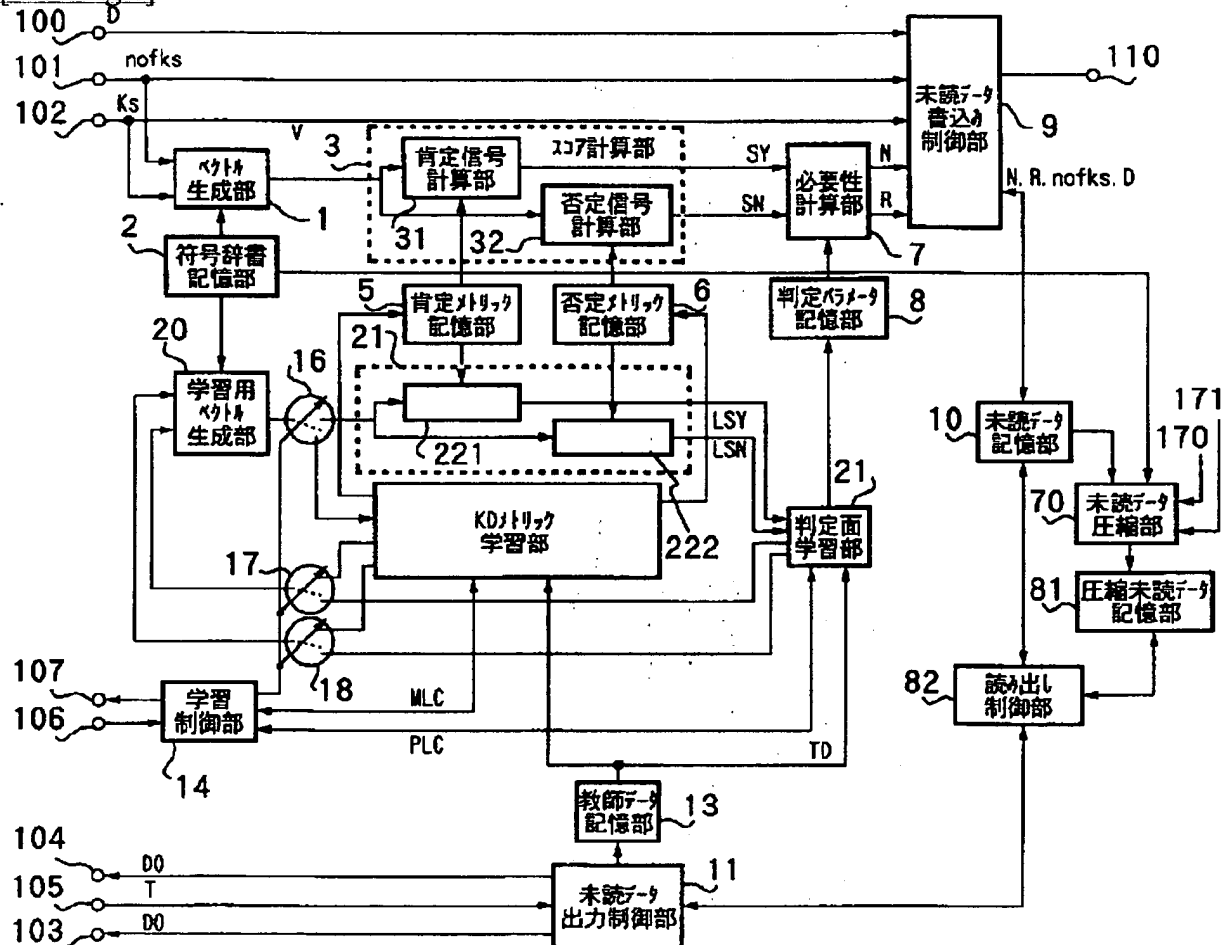
## \* NOTICES \*

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. \*\*\*\* shows the word which can not be translated.
3. In the drawings, any words are not translated.

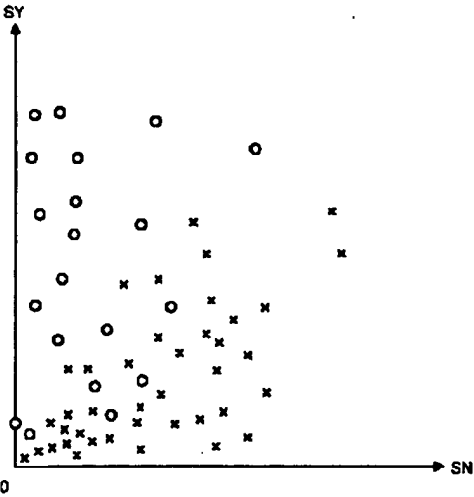
## DRAWINGS

[Drawing 1]

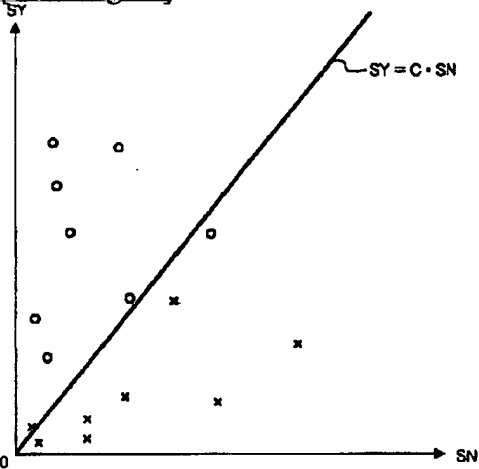


[Drawing 9]

○ T [C] = 1  
x T [C] = 0



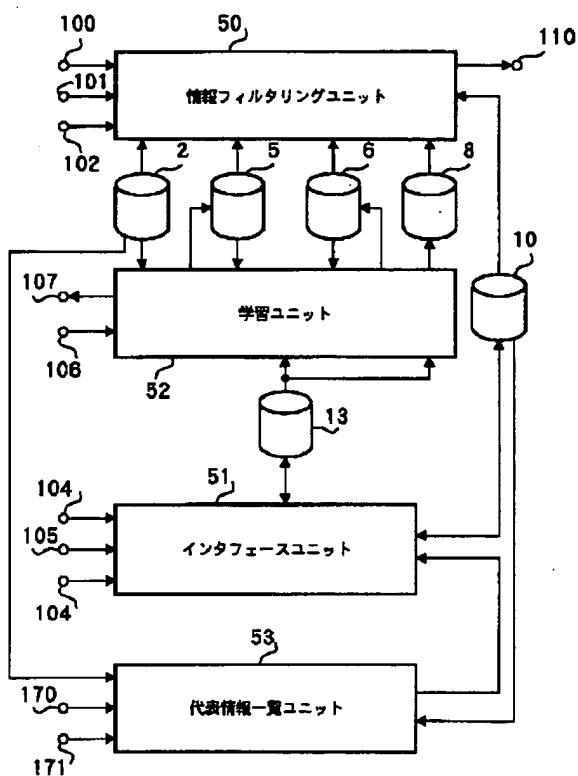
[Drawing 10]



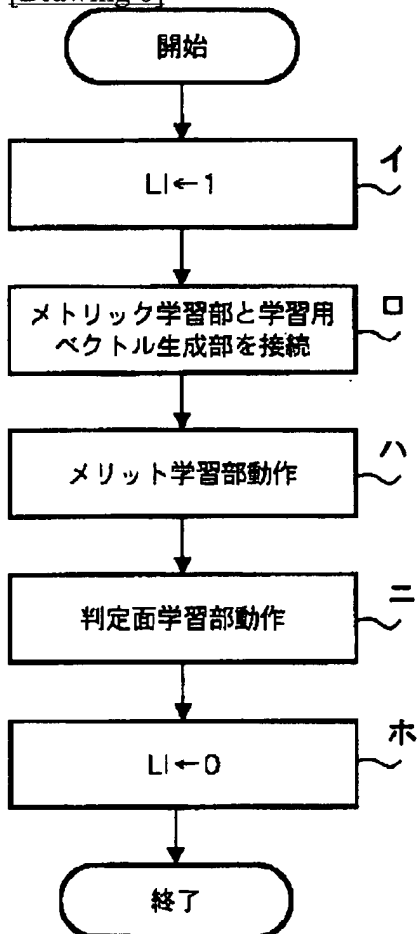
[Drawing 16]

1	3	5	-	-	-	-
2	-	-	-	-	-	-
4	6	7	-	-	-	-
8	10	-	-	-	-	-
9	-	-	-	-	-	-
<hr/>						
-	-	-	-	-	-	-

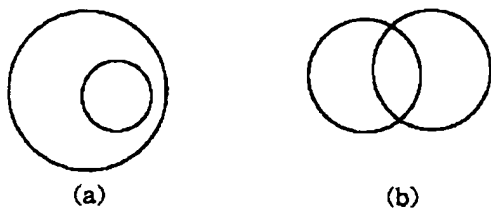
[Drawing 2]



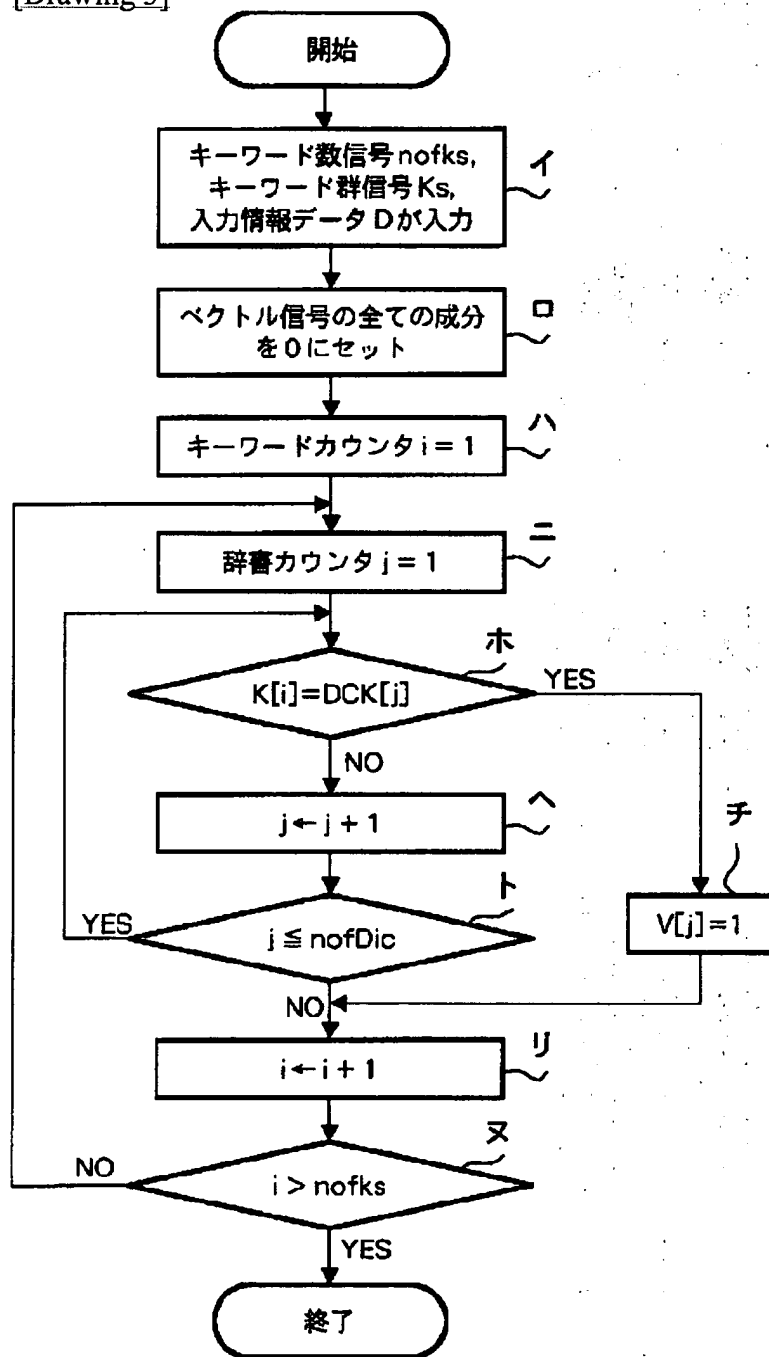
[Drawing 6]



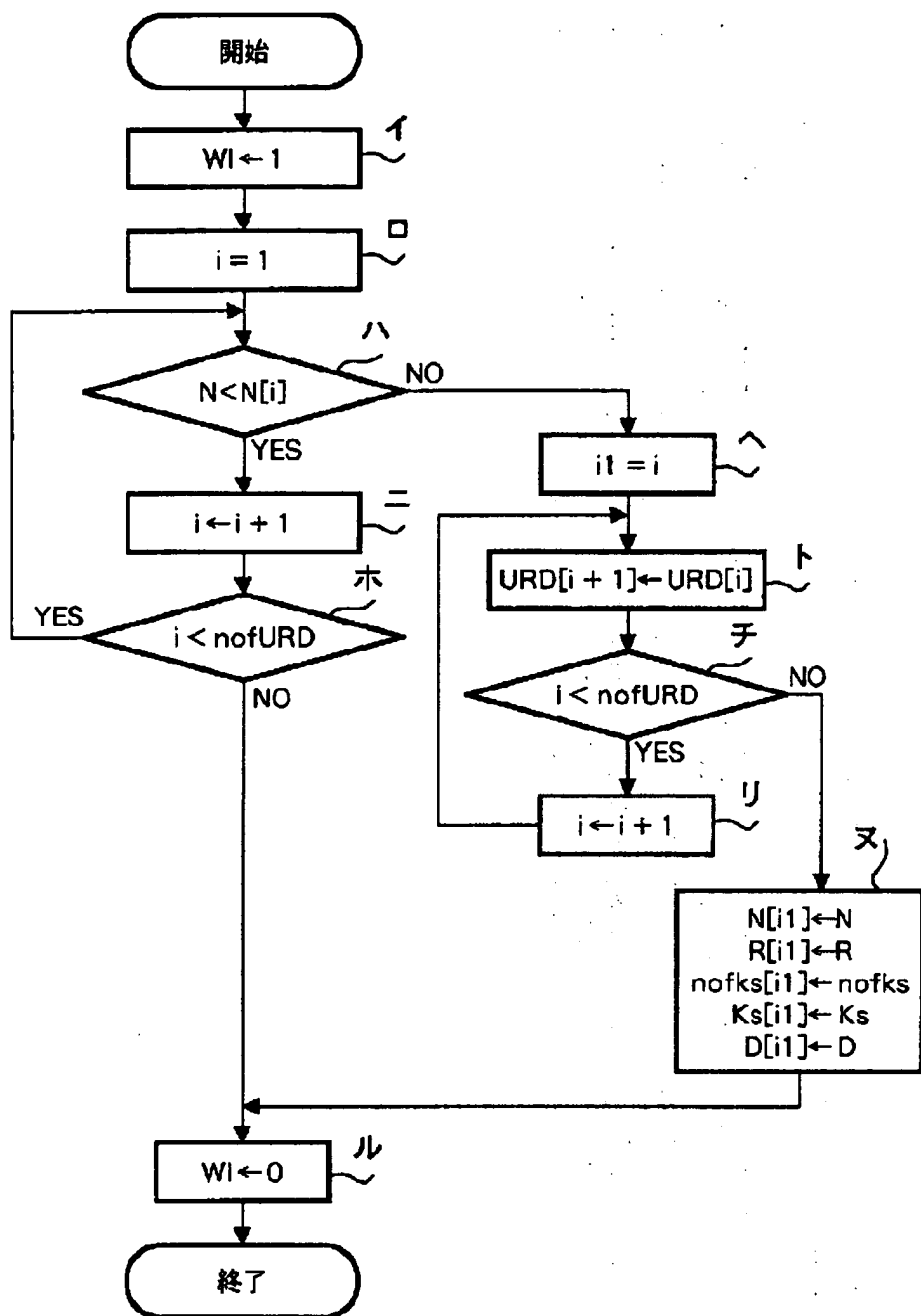
[Drawing 17]



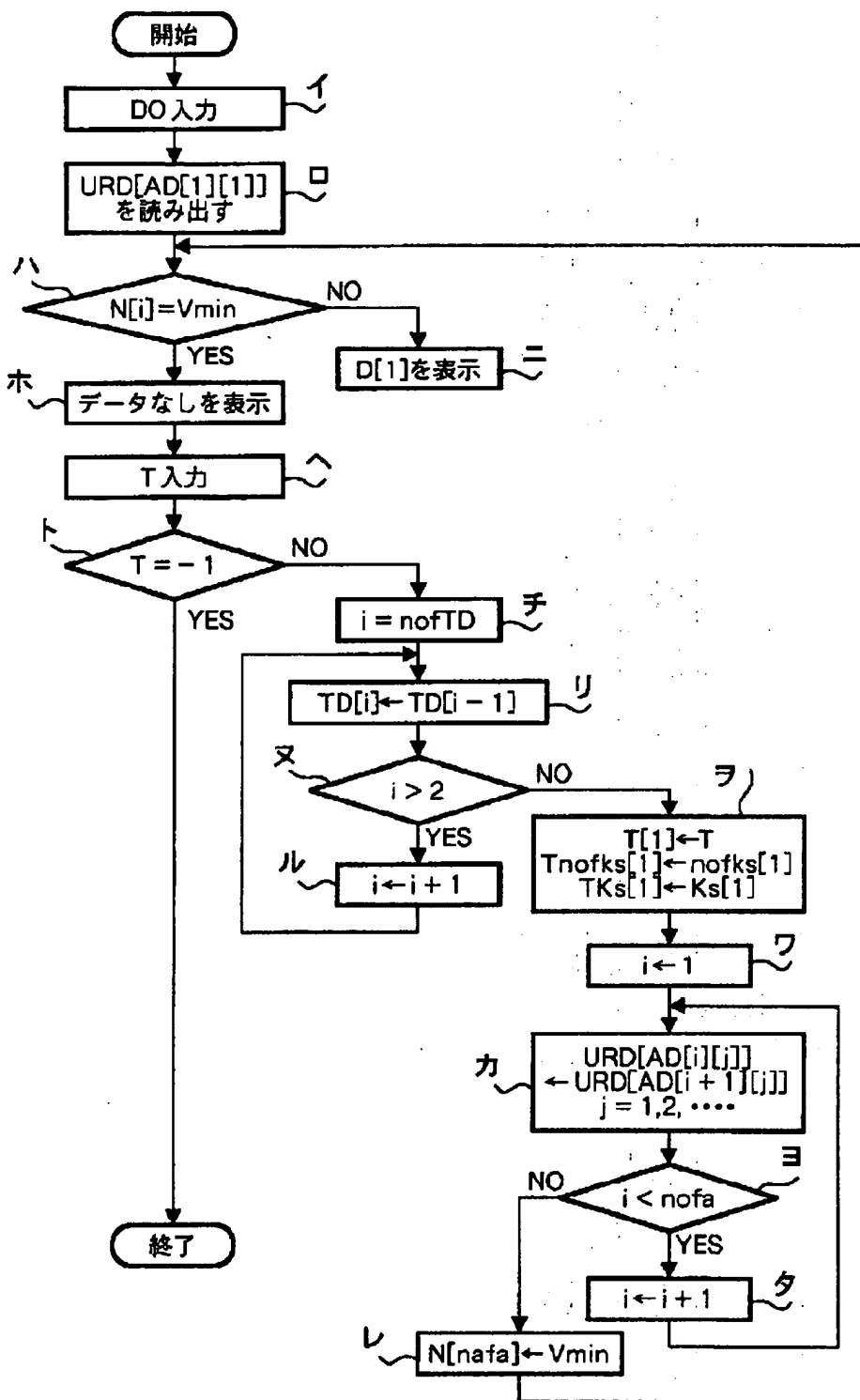
[Drawing 3]



[Drawing 4]

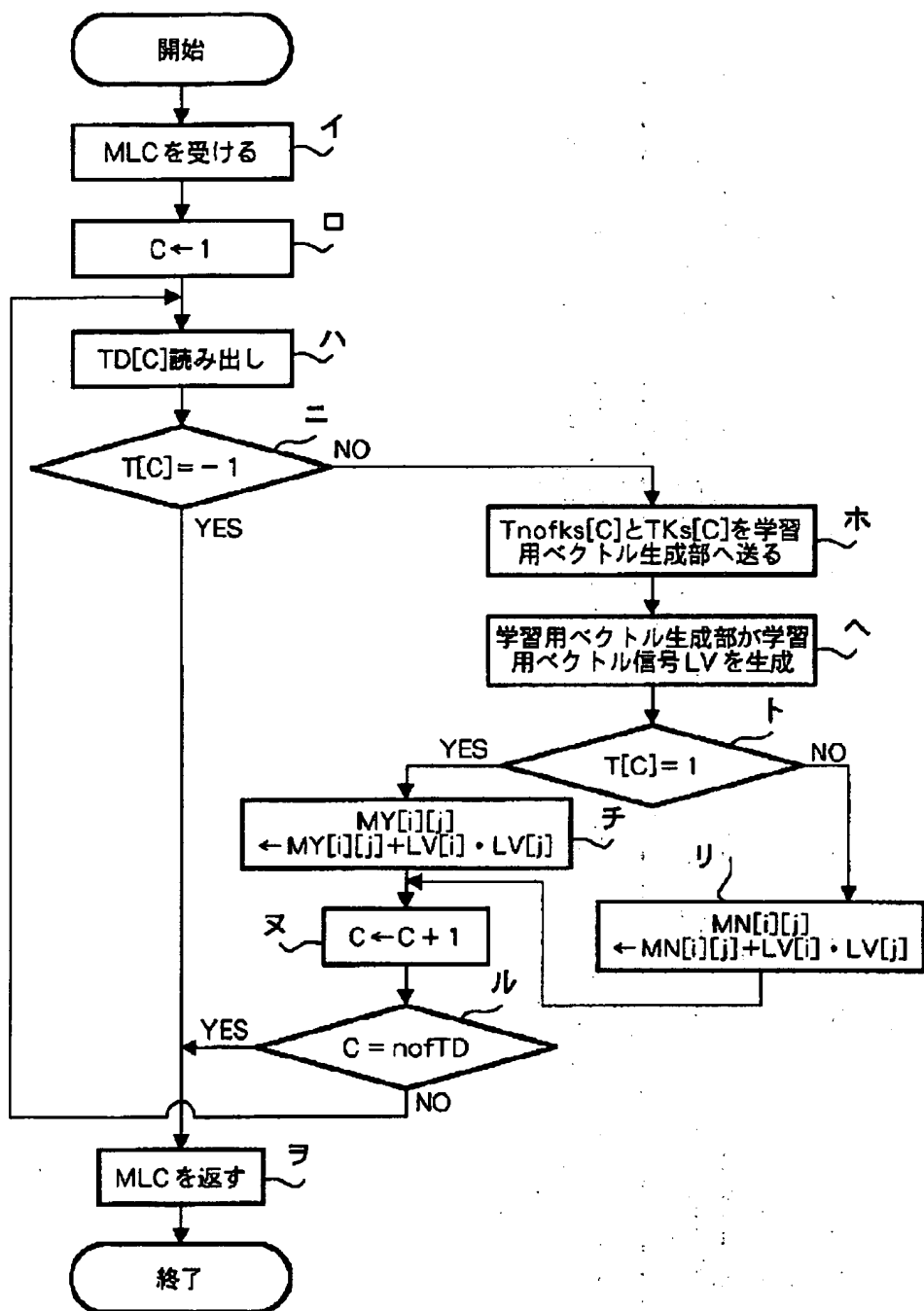


[Drawing 5]

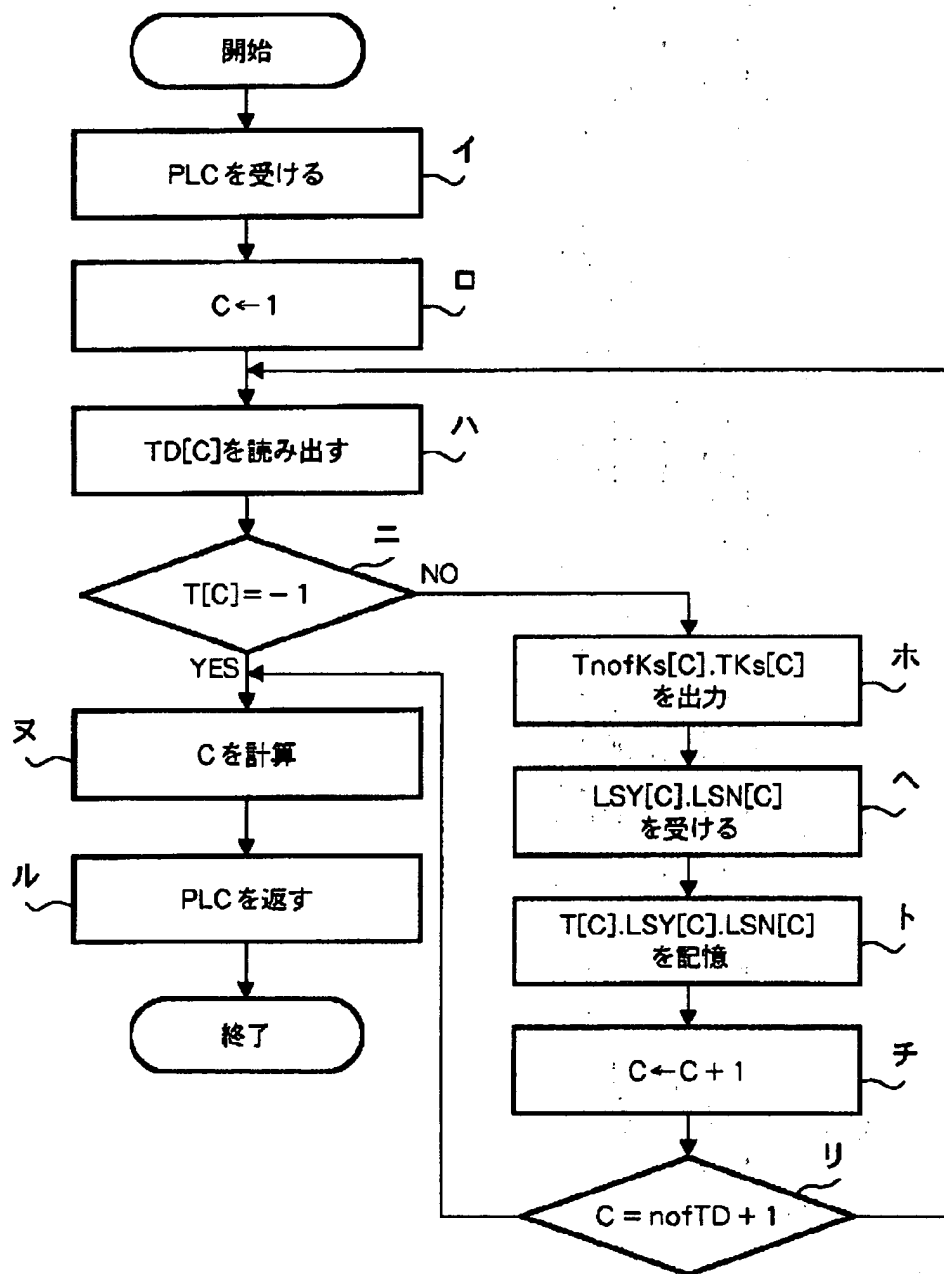


[Drawing 7]

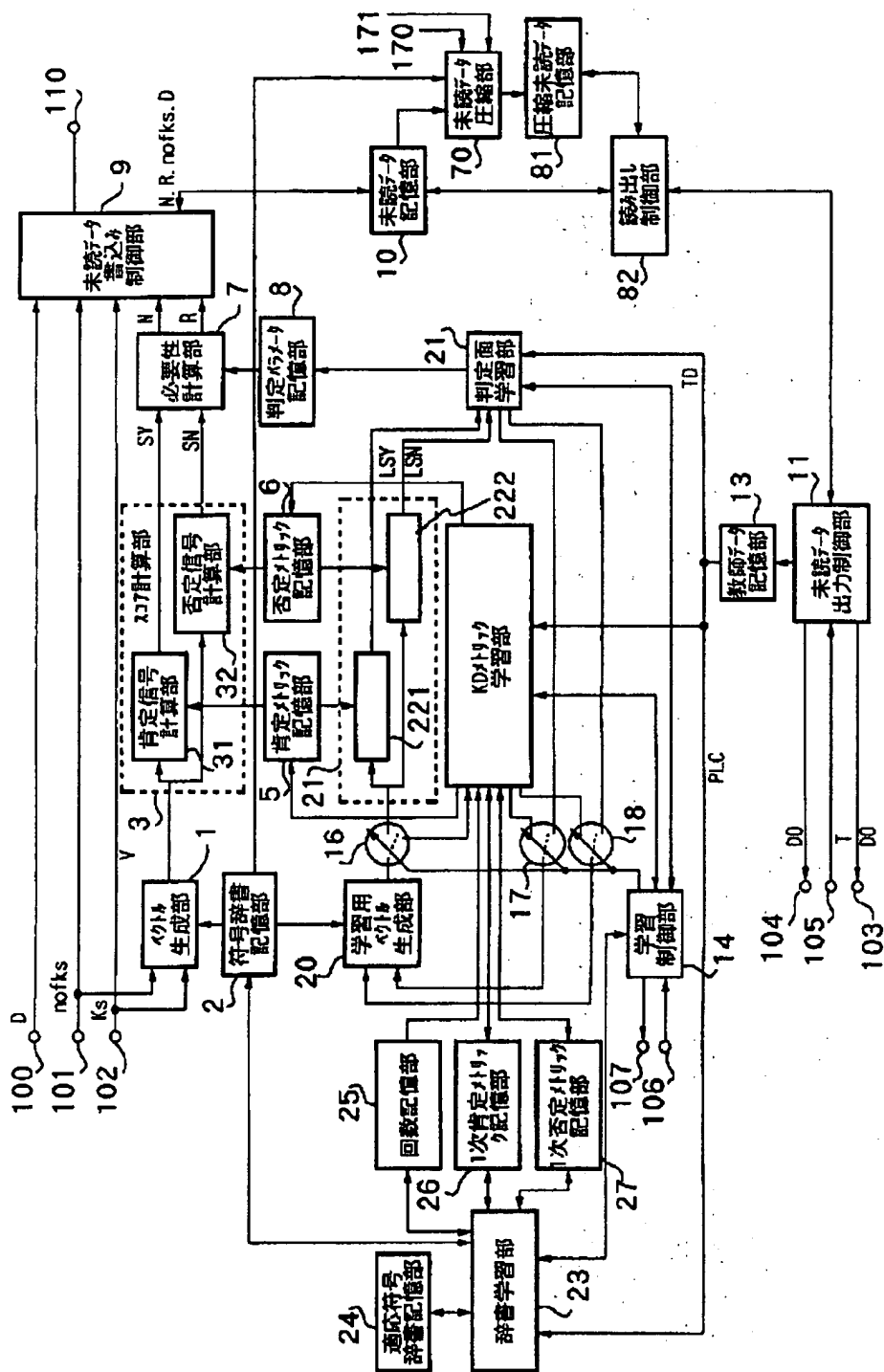




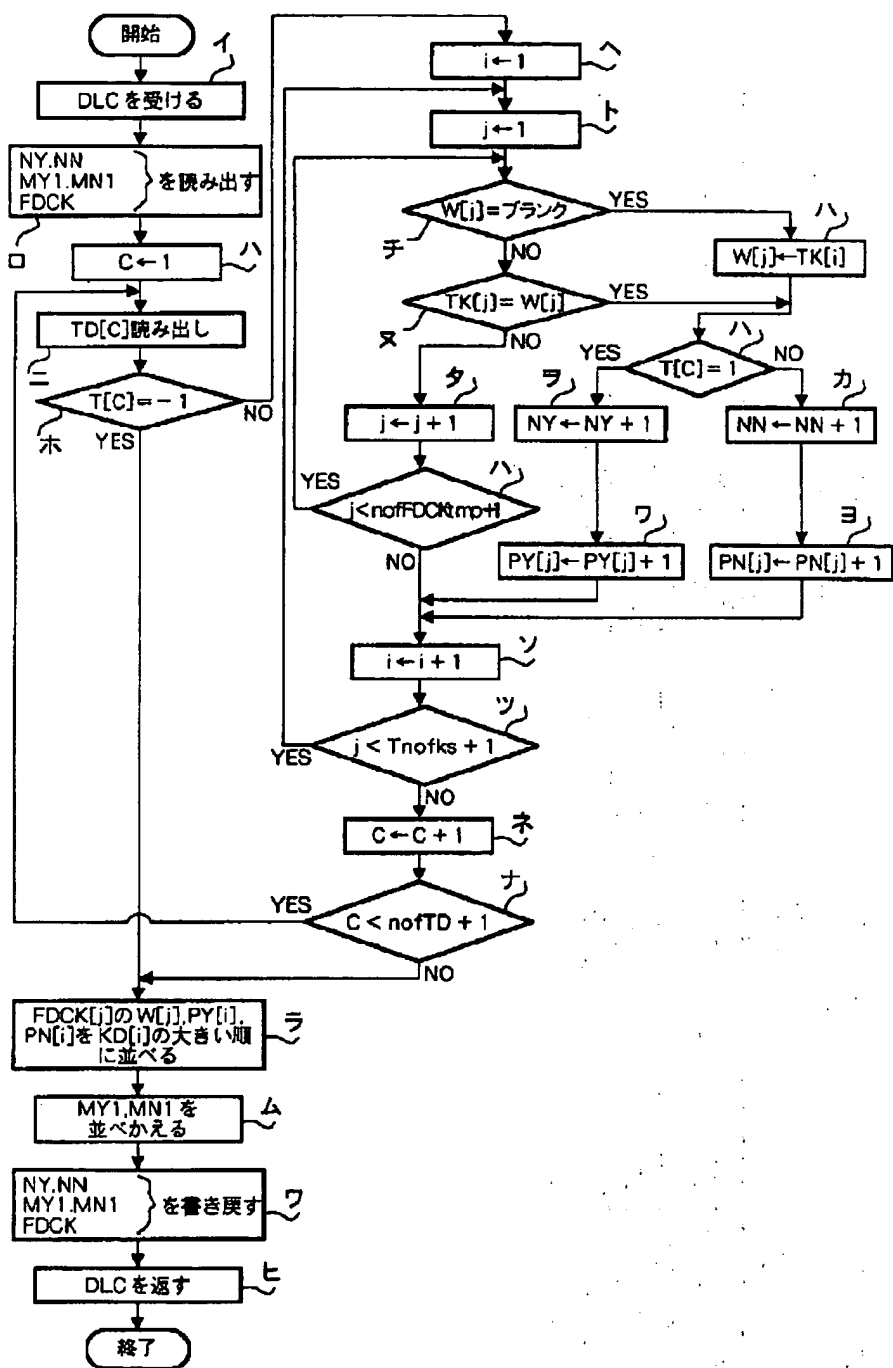
[Drawing 8]



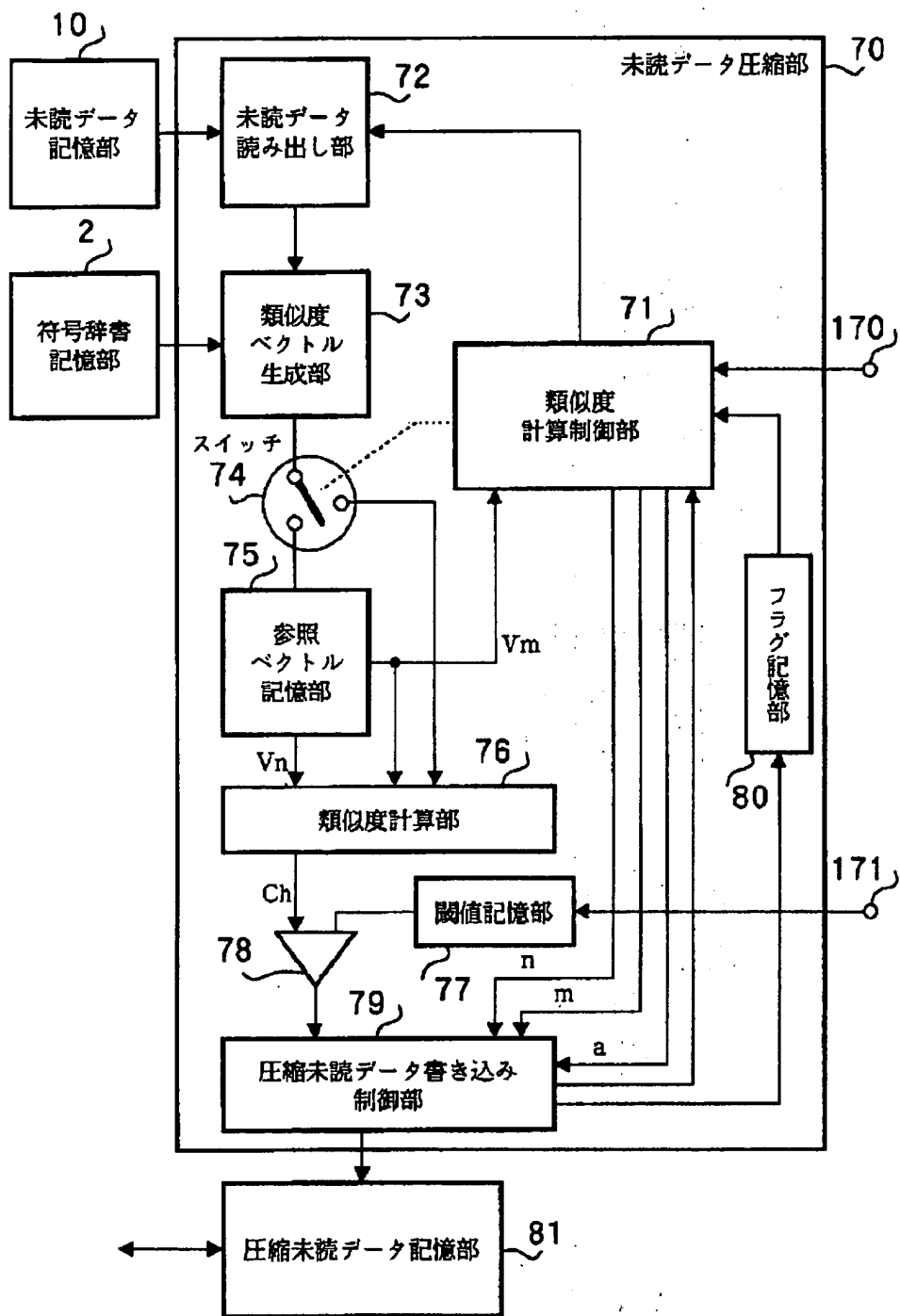
[Drawing 11]



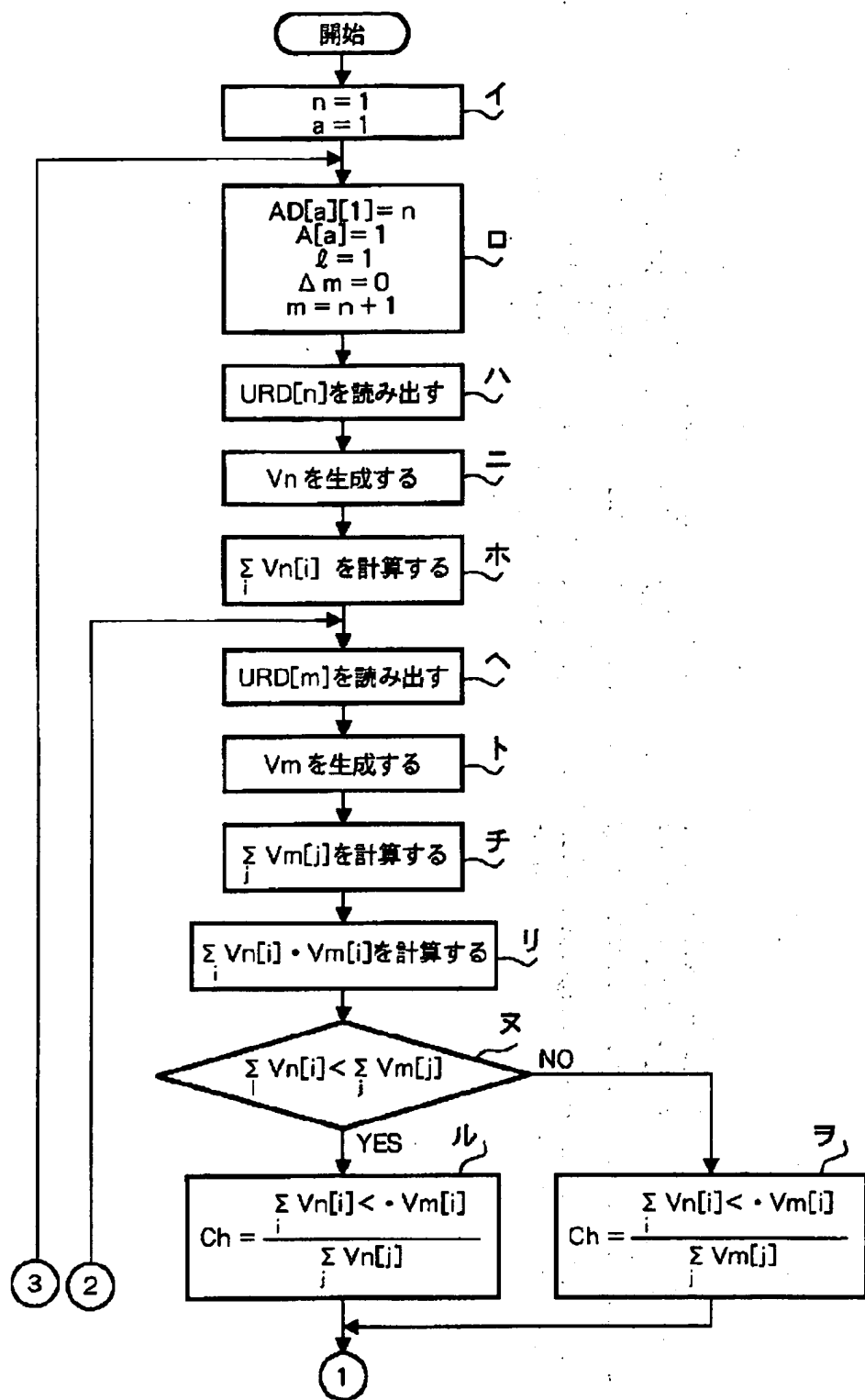
[Drawing 12]



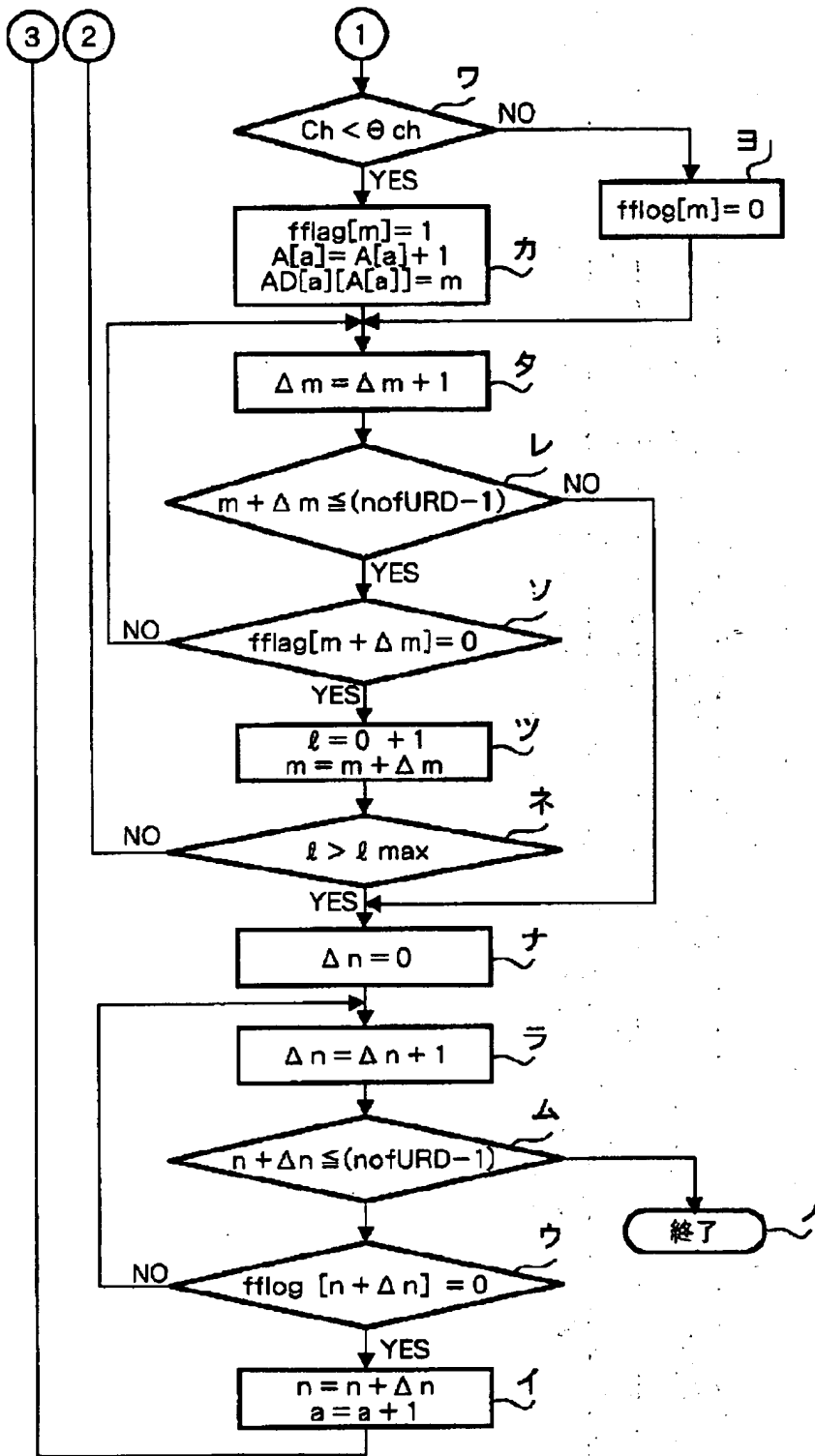
[Drawing 13]



[Drawing 14]



[Drawing 15]



[Translation done.]